


```

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB012003; AAM38389.1; -.
DR InterPro; IPR008640; Hep Hag.
DR 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DR 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAG
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=012E;
RC MEDLINE=22112901; PubMed=12117964;
RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Geme J.W. III,
RA Hansen E.J.;
RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
RT Activities.";
RL Infect. Immun. 70:4523-4533(2002).
DR EMBL; AY077638; AAL78285.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 9.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 2314 AA; 238336 MW; 3FFD721F67D4E2FD CRC64;

Query Match 42.9%; Score 84; DB 2; Length 2314;
Best Local Similarity 56.4%; Pred. No. 0.45;
Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

OY 5 SEADGGKGGANAR----GDKSIAIGDIAQAALGSQSIAIG 39
Db 2040 AKADGEAAVAIGRTQAGNQSIAGDQAATGDSIAIG 2078

RESULT 3
OY Q8PGS0 PRELIMINARY; PRT; 2190 AA.
AC Q8PGS0;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XAC3546.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RC MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cicarelli R.M.B., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RN Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB012164; AAM39974.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 59.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2351 AA; 220261 MW; C5482A38C940DAll CRC64;

Query Match 39.3%; Score 77; DB 16; Length 2351;
Best Local Similarity 60.7%; Pred. No. 3.2;
Matches 17; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 12 GGNARGDKSIAIGDIAQALGSQSIAG 39
Db 863 GOSNAAGDESIALGWEAQAEGDQGIAG 890

RESULT 5
Q9LAX0 PRELIMINARY; PRT; 1328 AA.
AC Q9LAX0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Xanthomonas campestris pv. pelargonii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=91612;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xpel-1;
RA Kim J.F., Zumoff C.H., Beer S.V.;
RT "An alanine-, glycine-, and serine-rich protein and a putative serine
RT protease of Xanthomonas campestris pv. pelargonii.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083618; AAF63394.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 26.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1328 AA; 127402 MW; ACE01A4C6548364C CRC64;

Query Match 36.5%; Score 71.5; DB 2; Length 1328;
Best Local Similarity 46.5%; Pred. No. 7.7;
Matches 20; Conservative 5; Mismatches 13; Indels 5; Gaps 2;

QY 3 GISEADG---GKGG--ANARGDKSIAIGDIAQALGSQSIAGD 40
Db 236 GLSTAGSLSSAGGYLSRASGDASTAFGYRARGSSSIAGVD 278

RESULT 6
Q9F3X6 PRELIMINARY; PRT; 1299 AA.
ID Q9F3X6;
AC Q9F3X6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MapA protein (Hsf).
GN MAPA OR HSF.2 OR PM1570.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AJ277635; CAC14202.1; -.
DR EMBL; AE006194; AAK03654.1; -.
DR InterPro; IPR008640; Hep_Hag.

```

```

DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 10.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB6CDB428 CRC64;

Query Match 36.2%; Score 71; DB 16; Length 1299;
Best Local Similarity 45.5%; Pred. No. 8.7;
Matches 20; Conservative 7; Mismatches 11; Indels 6; Gaps 2;

QY 2 IGISEADGKGG---GANARGDK--SIAIGDIAQALGSQSIAG 39
Db 109 IGFGATNDGETNVAIGAKSKSKAASIAIGDNAKALDQAIAIG 152

RESULT 7
Q83J23 PRELIMINARY; PRT; 382 AA.
ID Q83J23;
AC Q83J23;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Orf, hypothetical protein.
GN SF3640.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; AE015372; AAN45087.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 382 AA; 37760 MW; F686328578741C10 CRC64;

Query Match 34.4%; Score 67.5; DB 16; Length 382;
Best Local Similarity 33.3%; Pred. No. 6.1;
Matches 18; Conservative 8; Mismatches 9; Indels 19; Gaps 2;

QY 5 SEADG---GKGG-----GANARGDKSIAIGDIAQALGSQSIAG 39
Db 267 NKADGVDAITAGNGSQSRGLNTIALGTASNTGDKSLALGSSANGINSVALG 320

RESULT 8
Q8XDGA PRELIMINARY; PRT; 1588 AA.
ID Q8XDGA;
AC Q8XDGA;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin (Hypothetical protein).
GN Z5029 OR ECS4480 OR H161.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

```

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / RIMD 0509952;
 RC MEDLINE=21156231; PubMed=11258796;
 RX Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=RIMD 0509952;
 RC Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,
 RA Shinagawa H.;
 RT "O157 specific gene similar to H. influenzae adhesin gene."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005587; AAG58749.1; -;
 DR EMBL; AP002566; BAB37903.1; -;
 DR EMBL; AB036416; BAB87814.1; -;
 DR PIR; AB6036; AB6036.
 DR PIR; H91188; H91188.
 DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep_Hag; 13.
 DR Pfam; PF05662; HIM; 12.
 DR Pfam; PF03895; Yada; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1588 AA; 160150 MW; B2BA4E06BFF28DEC CRC64;
 Query Match 34.4%; Score 67.5; DB 16; Length 1588;
 Best Local Similarity 33.3%; Pred. No. 28;
 Matches 18; Conservative 8; Mismatches 9; Indels 19; Gaps 2;
 Qy 5 SEADG-----GKG-----GANARGDKSIAIGDIAQAALGSQSIAG 39
 Db 281 NKAGDVDAIALGNGSQSRGLNTIALGTASNATGDKSLALGSSNGAINSVALG 334
 RESULT 9
 Q89IU6
 ID Q89IU6 PRELIMINARY; PRT; 762 AA.
 AC Q89IU6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Br5538 protein.
 GN BLR5538.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobiium.
 OX NCBI_TaxID=375;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=USDA 110;
 RC MEDLINE=22484998; PubMed=12597275;
 RX Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iriuchii M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110."
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005955; BAC50803.1; -;

DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR Pfam; PF05658; Hep_Hag; 12.
 DR Pfam; PF05662; HIM; 3.
 KW Complete proteome.
 SQ SEQUENCE 762 AA; 72776 MW; B2DFEE9A1CB241CE CRC64;
 Query Match 34.2%; Score 67; DB 16; Length 762;
 Best Local Similarity 48.6%; Pred. No. 15;
 Matches 17; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 Qy 5 SEADGKGKANARGDKSIAIGDIAQAALGSQSIAG 39
 Db 521 TKVNSAGAAANASGTDIAIGNAQAATGSGSIAIG 555
 RESULT 10
 Q9FDA0
 ID Q9FDA0 PRELIMINARY; PRT; 1265 AA.
 AC Q9FDA0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative outer membrane protein Xada.
 GN XADA.
 OS Xanthomonas oryzae (pv. oryzae).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=64187;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Ray S.K., Rajeshwari R., Sonti R.V.;
 RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
 that is involved in virulence."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF288222; AAG01335.1; -;
 DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep_Hag; 25.
 DR Pfam; PF05662; HIM; 4.
 DR Pfam; PF03895; Yada; 1.
 SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;
 Query Match 34.2%; Score 67; DB 2; Length 1265;
 Best Local Similarity 46.2%; Pred. No. 25;
 Matches 12; Conservative 8; Mismatches 6; Indels 0; Gaps 0;
 Qy 14 ANARGDKSIAIGDIAQAALGSQSIAG 39
 Db 694 AQATGVSVAALGEISKATGESVAVG 719
 RESULT 11
 Q92KQ7
 ID Q92KQ7 PRELIMINARY; PRT; 1291 AA.
 AC Q92KQ7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein R00458.
 GN R00458 OR SMC01708.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=1021;
 RC MEDLINE=21396507; PubMed=11481430;
 RX Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,


```

RA Pohl T., Portetelle D., Puhler A., Fumelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591783; CAC41895.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR008640; Hep.Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep.Hag; 13.
DR Pfam; PF05662; HIM; 6.
DR Pfam; PF03895; Yada; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1291 AA; 127509 MW; 1B0F2A8CD1B1613C CRC64;

Query Match 34.2%; Score 67; DB 16; Length 1291;
Best Local Similarity 44.2%; Pred. No. 26;
Matches 19; Conservative 3; Mismatches 13; Indels 8; Gaps 1;

QY 5 SEADGGKGG-----ANARGKSIAGDIAQALGQSQTAI 39
DB 88 NEADFGASDATAIGTDAQNGSLAIGRQNGNEQSIGIG 130

RESULT 12
ID Q8FCB2 PRELIMINARY; PRT; 1778 AA.
AC Q8FCB2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin.
GN C4424.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CF7073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016768; AAN82860.1; -.
DR InterPro; IPR008640; Hep.Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep.Hag; 14.
DR Pfam; PF05662; HIM; 15.
DR Pfam; PF03895; Yada; 1.
DR KW Complete proteome.
SQ SEQUENCE 1778 AA; 177678 MW; 98564AA3A797DA20 CRC64;

Query Match 34.2%; Score 67; DB 16; Length 1778;
Best Local Similarity 39.6%; Pred. No. 37;
Matches 19; Conservative 6; Mismatches 11; Indels 12; Gaps 1;

QY 5 SEADGGKGGANRG-----DKSIAGDIAQALGQSQTAI 40
DB 217 SLAFGRKSTANSTGSLAIGADSSSDNDAIGNKTOALGVNSWALGN 264

RESULT 13
Q9EQT3
ID Q9EQT3 PRELIMINARY; PRT; 261 AA.

```

```

Q9EQT3;
AC 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GTP-binding protein like 1 (Wrch-1).
DR ARHU OR WRCH1 OR MG28K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Daigo Y., Takayama I., Fujino M.A.;
RT "Isolation, mapping, and characterization of a novel murine cDNA,
RT MG28K homologous to the mammalian GTP-binding protein family.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21352806; PubMed=11459829;
RA Tao W., Pennica D., Xu L., Kalejta R.F., Levine A.J.;
RT "Wrch-1, a novel member of the Rho gene family that is regulated by
RT Wnt-1.";
RL Genes Dev. 15:1796-1807(2001).
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
DR EMBL; AB051827; BAB18639.1; -.
DR EMBL; AF378088; AAK83341.1; -.
DR HSSP; P21181; 1AM4.
DR MGD; MGI:1316831; Arhu.
DR GO; GO:0003931; F:Rho small monomeric GTPase activity; IMP.
DR GO; GO:0003036; P:Actin cytoskeleton organization and biogenesis; IDA.
DR GO; GO:0000082; P:G1/S transition of mitotic cell cycle; IDA.
DR GO; GO:0008360; P:regulation of cell shape; IDA.
DR InterPro; IPR003578; GTPase_Rho.
DR InterPro; IPR001806; Ras_trnsmng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00174; RHO; 1.
DR TIGRFAMS; TIGR00231; small GTP; 1.
DR KW GTP-binding; Lipoprotein; Prenylation.
SQ SEQUENCE 261 AA; 28353 MW; 8EC4250071E75B14 CRC64;

Query Match 33.7%; Score 66; DB 11; Length 261;
Best Local Similarity 39.5%; Pred. No. 6.1;
Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 1;

QY 1 GTGISADGGKGGANRGDKSIAGDIAQALGQSQTAI 38
DB 35 GPGVSGRGRAGGAGRGVGVKVLVGD--GAVGKTSLVV 70

RESULT 14
ID Q9PD63 PRELIMINARY; PRT; 1004 AA.
AC Q9PD63;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Surface-exposed outer membrane protein.
DR XFI516.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colaudo N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:32:39 ; Search time 0.743764 Seconds
(without alignments)
2800.358 Million cell updates/sec

Title: US-09-813-214A-11

Perfect score: 196

Sequence: 1 GIGISEADGGKGNARGDKSIAGIDIAQAALGQSQAIGD 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	31.1	377	1 RPOA_CHLMU	Q9p1n4 Chlamydia m
2	61	31.1	455	2 YADA_YEREN	P31489 yersinia en
3	60	30.6	278	1 T2D7_DROME	Q27272 drosophila
4	59.5	30.4	434	1 YADA_YERPS	P10858 yersinia ps
5	58.5	29.8	104	1 CH10_SCHPO	O59804 schizosach
6	57.5	29.3	403	1 CK05_HUMAN	Q9uhr6 homo sapien
7	57.5	29.3	676	1 ICPO_HSVBK	P29836 bovine herp
8	57.5	29.3	710	1 EFG_BUCBP	P59451 pseudomonas
9	57.5	29.3	725	1 MASZ_PSEFL	O05137 buchnera ap
10	57	29.1	1901	1 YZ08_MYCTU	O53553 mycobacteri
11	57	29.1	2946	1 NBEA_HUMAN	Q8nfp9 homo sapien
12	56	28.6	377	1 RPOA_CHLTR	Q46449 chlamydia t
13	55.5	28.3	498	1 NDDD_ALCXX	P94212 alcaligenes
14	55.5	28.3	698	1 EFG_VIBCH	Q9kuz7 vibrio chol
15	54.5	27.8	258	1 BDHA_ALCEU	O9x6u2 alcaligenes
16	54.5	27.8	730	1 KOGA_MOUSE	O88673 mus musculu
17	54	27.6	245	1 YH77_ARCFU	O28497 archaeoglob
18	54	27.6	370	1 LEU3_BRAJA	Q89x14 bradyrhizob
19	53.5	27.3	331	1 MACS_BOVIN	P12624 bos taurus
20	53.5	27.3	699	1 EFG_HABIN	P43925 haemophilus
21	53.5	27.3	699	1 EFG_VIBPA	Q87145 vibrio para
22	53.5	27.3	699	1 EFG_VIBVU	Q8dgc8 vibrio vuln
23	53.5	27.3	700	1 EFG_PASMU	P57938 pasteurella
24	53.5	27.3	703	1 EFG_ECOLI	P02996 escherichia
25	53.5	27.3	703	1 EFG_SALTY	P26229 salmonella
26	53.5	27.3	734	1 KOGA_PIG	P20192 sus scrofa
27	53	27.0	219	1 PGMB_ECOLI	P77366 escherichia
28	53	27.0	401	1 YK03_CABEL	P34291 caenorhabdi
29	53	27.0	555	1 CSW_DROSI	Q8isc9 drosophila
30	53	27.0	595	1 PRIM_CHLTR	O84799 chlamydia t
31	52.5	26.8	329	1 HEM2_MYCLE	P46723 mycobacteri
32	52.5	26.8	459	1 COBG_PSEDE	P21637 pseudomonas
33	52.5	26.8	698	1 EFG1_SHEON	Q8ek71 shewanella

RESULT 1
RPOA_CHLMU
ID_RPOA_CHLMU STANDARD; PRT; 377 AA.
AC Q9PJN4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR TC0794.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Ni9g;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback I., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";
RT Nucleic Acids Res. 28:1397-1406(2000).
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
CC -!- SUBUNIT: Homodimer. The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta', and 1 omega subunit (By similarity).
CC -!- DOMAIN: The N-terminal domain is essential for RNAP assembly and basal transcription, whereas the C-terminal domain is involved in interaction with transcriptional regulators and with upstream promoter elements (By similarity).
CC -!- SIMILARITY: Belongs to the RNA polymerase alpha chain family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AR002347; AAF39597.1; ALT_INIT.
HSSP; P00574; 1COO.
TIGR; TC0794; -.
HAMAP; MF 00059; -; 1.
InterPro; IPR009025; RBP11-like RNAPo.
InterPro; IPR001700; RNA_pol_A_bac_org.
Pfam; PF01000; RNA_pol_A_bac_1.
Pfam; PF03118; RNA_pol_A_CTD; 1.
ProDom; PD001179; RNA_pol_A_bac_org; 1.
SMART; SM00662; RPOLD; 1.

Q83es7 coxiella bu
P57593 buchnera ap
Q8k948 buchnera ap
Q8zjb3 yersinia pe
Q88fi4 pseudomonas
Q8d3h2 wiggleswort
P51556 rattus norv
Q8fsv1 corynebacte
Q971w0 sulfolobus
Q88wb6 lactobacill
Q9rn37 salmonella
P06196 salmonella

KW Transferase; Transcription; DNA-directed RNA polymerase;
 KW Complete proteome.
 FT DOMAIN 1 259 ALPHA N-TERMINAL DOMAIN (ALPHA-NTD)
 (BY SIMILARITY).
 FT DOMAIN 276 377 ALPHA C-TERMINAL DOMAIN (ALPHA-CTD)
 (BY SIMILARITY).
 FT SEQUENCE 377 AA; 41833 MW; 7DB20C7206EBD2E6 CRC64;
 SQ
 Query Match 31.1%; Score 61; DB 1; Length 377;
 Best Local Similarity 32.4%; Pred. No. 6.1;
 Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 QY 4 ISEADGKGKGNARGDKSIAIGDIAQALGSSQSIAGD 40
 : : : : :
 DB 106 LQDCEGGRASQKLRATISIDASDLAAAGGQKAITLGD 142
 : : : : :
 RESULT 2
 YADA_YEREN
 ID YADA_YEREN STANDARD; PRT; 455 AA.
 AC P31489;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Invasin precursor (Outer membrane adhesion).
 GN YADA OR YOPA OR INVA OR YOP1.
 OS Yersinia enterocolitica.
 OG Plasmid pYV.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6471/76 / Serotype O:3;
 RX MEDLINE=95020586; PubMed=7934875;
 RA Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,
 RA Skurnik M.;
 RT "Hydrophobic domains affect the collagen-binding specificity and
 RT surface polymerization as well as the virulence potential of the Yada
 RT protein of Yersinia enterocolitica.";
 RL Mol. Microbiol. 10:995-1011(1993).
 CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
 CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
 CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
 CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
 CC SURFACE.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X13882; CAA32086.1; -.
 DR FIR; S04912; S04912.
 DR InterPro; IPR008126; Adhesion.
 DR InterPro; IPR008640; Hep.Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep.Hag; 4.
 DR Pfam; PF05662; HIM; 1.
 DR Pfam; PF03895; Yada; 1.
 DR PRINTS; PR01756; OMADHESIN.
 DR Plasmid; Virulence; signal; Outer membrane.
 FT SIGNAL 1 25
 FT CHAIN 26 455 INVASIN.
 SQ SEQUENCE 455 AA; 47136 MW; AC12EF68C657DAC0 CRC64;
 Query Match 31.1%; Score 61; DB 1; Length 455;
 Best Local Similarity 46.9%; Pred. No. 7.4;
 KW
 KW Complete proteome.
 FT DOMAIN 1 259 ALPHA N-TERMINAL DOMAIN (ALPHA-NTD)
 (BY SIMILARITY).
 FT DOMAIN 276 377 ALPHA C-TERMINAL DOMAIN (ALPHA-CTD)
 (BY SIMILARITY).
 FT SEQUENCE 377 AA; 41833 MW; 7DB20C7206EBD2E6 CRC64;
 SQ

Matches 15; Conservative 6; Mismatches 9; Indels 2; Gaps 1;
 QY 10 GKGG-ANARGDKSIAIGDIAQALGSSQSIAG 39
 : : : : :
 DB 59 GAGGINASAKGHSIAIGATAEAAGAAGAAVAVG 90
 : : : : :
 RESULT 3
 T2D7_DROME
 ID T2D7_DROME STANDARD; PRT; 278 AA.
 AC Q27272; Q9VK16;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transcription initiation factor TFIID 42 kDa subunit (TAFII-42)
 DE (TAFII40) (p42) (Enhancer of yellow 1 protein).
 GN E(Y)1 OR TAF40 OR CG6474.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150630; PubMed=7545910;
 RA Kokubo T., Gong D.W., Wootton J.C., Horikoshi M., Roeder R.G.,
 RA Nakatani Y.;
 RT "Molecular cloning of Drosophila TFIID subunits.";
 RL Nature 367:484-487(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=94037099; PubMed=8221891;
 RA Goodrich J.A., Hoey T., Thut C.J., Admon A., Tjian R.;
 RT "Drosophila TAFII40 interacts with both a VP16 activation domain and
 RT the basal transcription factor TFIIB.";
 RL Cell 75:519-530(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Vardell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flesler C., Gabrielian A.E., Garg N.S., Galbraith W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: TAFs are components of the transcription factor IID
 CC (TFIID) complex that are essential for mediating regulation of RNA
 CC polymerase transcription (by similarity).
 CC -!- SUBUNIT: TFIID is composed of TATA binding protein (TBP) and a
 CC number of TBP-associated factors (TAFs).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the TAF2G family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: U06458; AAC47347.1; -.
 CC PIR: A49067; A49067.
 CC TRANSFAC: T02125; -.
 CC FlyBase: FBgn000617; e(y)1.
 CC InterPro: IPR007124; Hist.TAF.
 CC InterPro: IPR003162; TFIID-31.
 CC Pfam: PF02291; TFIID-31; 1.
 CC ProDom: PD011023; TFIID-31; 1.
 CC Transcription regulation; Nuclear protein.
 CC DOMAIN 190 197 POLY-THR.
 CC DOMAIN 202 209 POLY-GLY.
 CC DOMAIN 250 256 POLY-GLY.
 CC DOMAIN 267 274 POLY-GLU (ACIDIC).
 CC SEQUENCE 278 AA; 29314 MW; 0EA442C80467001F CRC64;
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 190 197 POLY-THR.
 FT DOMAIN 202 209 POLY-GLY.
 FT DOMAIN 250 256 POLY-GLY.
 FT DOMAIN 267 274 POLY-GLU (ACIDIC).
 SQ SEQUENCE 278 AA; 29314 MW; 0EA442C80467001F CRC64;
 Query Match 30.6%; Score 60; DB 1; Length 278;
 Best Local Similarity 31.6%; Pred. No. 5.9;
 Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;
 QY 2 IGISEADGGKGNARGDKSIAIGDIAQALGSQSIATG 39
 Db 198 VGSSEGGSGGGQGVKSESTGAGGDLKMEVDSDAAVG 235
 RESULT 4
 YADA YERPS
 ID YADA YERPS STANDARD; PRT; 434 AA.
 AC P10858;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last annotation update)
 DE Invasin precursor (Outer membrane adhesin).
 GN YADA OR YOPA OR INVA OR YOP1.
 OS *Yersinia pseudotuberculosis*.
 OG Plasmid pIB1.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Yersinia*.
 OX NCBI_TaxID=633;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=YPIII;
 RX MEDLINE=88302441; PubMed=3043229;
 RA Rosqvist R., Skurnik M., Wolf-Watz H.;
 RT "Increased virulence of *Yersinia pseudotuberculosis* by two
 RT independent mutations.";
 RL Nature 334:522-525(1998).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=YPIII;

RX MEDLINE=89343638; PubMed=2761389;
 RA Skurnik M., Wolf-Watz H.;
 RT "Analysis of the yopA gene encoding the YopI virulence determinants
 of *Yersinia* spp.";
 RL Mol. Microbiol. 3:517-529(1989).
 CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
 CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
 CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
 CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
 CC SURFACE.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: X13983; CAA32088.1; -.
 CC PIR: S04534; S04534.
 CC InterPro: IPR008126; Adhesion.
 CC InterPro: IPR008640; Hep Hag.
 CC InterPro: IPR008635; HIM.
 CC InterPro: IPR005594; Yada.
 CC Pfam: PF05658; Hep Hag; 4.
 CC Pfam: PF05662; HIM; 1.
 CC Pfam: PF03895; Yada; 1.
 CC PRINTS: PR01756; OMADHESIN.
 KW Plasmid; Virulence; Signal; Outer membrane.
 FT SIGNAL 1 25
 FT CHAIN 26 434 INVASIN
 FT SEQUENCE 434 AA; 45054 MW; EE2C55FB12B183D4 CRC64;
 SQ
 Query Match 30.4%; Score 59.5; DB 1; Length 434;
 Best Local Similarity 48.7%; Pred. No. 10;
 Matches 19; Conservative 2; Mismatches 9; Indels 9; Gaps 2;
 QY 10 GKGGANARGDK--STAIGDIAQ-----ALGSQSIATG 39
 Db 94 GAGGLNARAKDPYSIAIGATAEAKPAAVAVGSGSIATG 132
 RESULT 5
 CH10 SCHPO
 ID CH10 SCHPO STANDARD; PRT; 104 AA.
 AC O59804;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 10 kDa heat shock protein, mitochondrial (HSP10) (10 kDa chaperonin).
 GN HSP10 OR SPCC550.06C.
 OS *Schizosaccharomyces pombe* (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC *Schizosaccharomyces*.
 OX NCBI_TaxID=4896;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Murgall K., Murphy L., Niblett D., Odell J.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,

DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MTW023.15.
OC Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL020222; CAA17745.1; .
DR PIR; F70806; F70806.
DR TubercuList; RV3508;
DR InterPro; IPR000084; PE region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE region; 1.
DR Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match 29.1%; Score 57; DB 1; Length 1901;
Best Local Similarity 59.1%; Pred. No. 85;
Matches 13; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
QY 1 GIGISEADGGKGANARGDKSI 22
||| : ||||| : |||
Db 1307 GIGGTGGAGGAGGAGDPSI 1328
RESULT 11
ID NBEA_HUMAN STANDARD; PRT; 2946 AA.
AC Q9NF99; Q9HCW8; Q9NSU1; Q9NW98; Q9Y6J1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurobeachin protein (lysosomal trafficking regulator 2) (BCL8B
protein).
DE NBEA OR LYST2 OR BCL8B OR KIAA1544.
GN NBEA OR LYST2 OR BCL8B OR KIAA1544.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RP TISSUE=Brain, and Spleen;
RC MEDLINE=22150869; PubMed=21160729;
RA Dymov V.G., Chaganti S.R., Dymova K., Palanisamy N., Murty V.V.V.S.,
RA Dalla-Favera R., Chaganti R.S.K.;

"BCL8 is a novel, evolutionarily conserved human gene family encoding
proteins with presumptive protein kinase A anchoring function.";
Genomics 80:158-165(2002).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RA Duesterhoeft A., Lauber J., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 606-1118 FROM N.A.
RC TISSUE=Embryonic head;
RA Itogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1919-2946 FROM N.A. (ISOFORM 1).
RX MEDLINE=20450683; PubMed=10957877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
XVIII. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
RN [5]
RP SEQUENCE OF 2428-2946 FROM N.A. (ISOFORM 1).
RA Tchernev V.T., McMurtre E.B., Nguyen Q.A., Mishra V.S.,
RA Barbosa M.D.F.S., McIndoe R., Kingmore S.F.;
RT "Identification of LYST2, a brain-specific member of the Chediak-
Higashi syndrome gene family.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 2150-2563.
RX MEDLINE=22220051; PubMed=12234919;
RA Jogi G., Shen Y., Gebauer D., Li J., Wiegmann K., Kashkar H.,
RA Krenke M., Tong L.;
RT "Crystal structure of the BRACH domain reveals an unusual fold and
extensive association with a novel PH domain.";
RL EMBO J. 21:4785-4795(2002).
CC -1- FUNCTION: Binds to type II regulatory subunits of protein kinase A
and anchors/targets them to the membrane. May anchor the kinase to
cytoskeletal and/or organelle-associated proteins (By
similarity).
CC -1- SUBUNIT: Interacts with RII subunit of PKA (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By
similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8NPP9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NPP9-2; Sequence=VSP_050538, VSP_050539;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Predominant in many brain structures. Also
expressed at medium levels in spleen, thymus, prostate, testis and
ovary. Low level expression is seen in heart, kidney, pancreas,
skeletal muscle and intestine.
CC -1- DOMAIN: RII-alpha binding site, predicted to form an amphipathic
helix, could participate in protein-protein interactions with a
complementary surface on the R-subunit dimer (By
similarity).
CC -1- SIMILARITY: Belongs to the neurobeachin family.
CC -1- SIMILARITY: Contains 1 BEACH domain.
CC -1- SIMILARITY: Contains 5 WD repeats.
CC -1- CAUTION: Ref.3 sequence differs from that shown due to a stop
codon in position 762.
CC -1- CAUTION: Ref.5 sequence differs from that shown due to a
frameshift in position 2900.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

AC P94212;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE N-acyl-D-aspartate decarboxylase (EC 3.5.1.83) (N-acyl-D-aspartate
 amidohydrolase).
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Achromobacter.
 OX NCBI_TaxID=515;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A-6;
 RA Wakayama M., Watanabe E., Takenaka Y., Miyamoto Y., Tau Y., Sakai K.,
 RA Moriguchi M.;
 RT "Cloning, expression, and nucleotide sequence of the N-acyl-D-
 RT aspartate amidohydrolase gene from Alcaligenes xylosoxydans subsp.
 RT xylosoxydans A-6.";
 RL J. Ferment. Bioeng. 80:311-317(1995).
 RN [2]
 RN CHARACTERIZATION.
 RC STRAIN=A-6;
 RX MEDLINE=93372486; PubMed=7763985;
 RA Moriguchi M., Sakai K., Katsuno Y., Maki T., Wakayama M.;
 RT "Purification and characterization of novel N-acyl-D-aspartate
 RT amidohydrolase from Alcaligenes xylosoxydans subsp. xylosoxydans
 RT A-6.";
 RL Biosci. Biotechnol. Biochem. 57:1145-1148(1993).
 CC -!- CATALYTIC ACTIVITY: N-acyl-D-aspartate + H(2)O = carboxylate + D-
 CC aspartate.
 CC -!- COFACTOR: Zinc.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the N-acyl-D-amino-acid deacylase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D45919; BAA08350.1; -;
 DR InterPro; IPR006680; Amidohydro.1.
 DR InterPro; IPR005847; Pept M38 regn.
 DR Pfam; PF01979; Amidohydro.1; 1.
 DR ProDom; PD000518; Urease; 1.
 DR Hydrolase; Zinc.
 KW SEQUENCE 498 AA; 53584 MW; EAF8662891F9F1B CRC64;
 SQ
 Query Match 28.3%; Score 55.5; DB 1; Length 498;
 Best Local Similarity 40.0%; Pred. No. 34;
 Matches 14; Conservative 4; Mismatches 12; Indels 5; Gaps 1;
 QY 3 GISEADCGGGAN-----ARGDKSIAIGDIAQLG 32
 DB 20 GATLDGGGPGARQGLAVRGRGRVALGDFAHAPG 54

RESULT 14
 EFG_VIBCH STANDARD; PRT; 698 AA.
 AC Q9KUZ7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUSA OR VC0361.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=E1 Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberger J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
 CC the nascent protein chain from the A-site to the P-site of the
 CC ribosome.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
 CC EF-G/EF-2 subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; A5004124; AAP93534.1; -;
 DR PIR; C82332; C82332.
 DR HSSP; P13551; LELO.
 DR TIGR; VC0361; -;
 DR HAMAP; MF 00054; -; 1.
 DR InterPro; IPR004540; EF-G.
 DR InterPro; IPR000795; EF GTPbind.
 DR InterPro; IPR000640; EFG_C.
 DR InterPro; IPR009022; EFG_III V.
 DR InterPro; IPR005517; EFG_IV.
 DR InterPro; IPR004161; EFTU D2.
 DR InterPro; IPR005225; Small GTP.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF00679; EFG_C; 1.
 DR Pfam; PF03764; EFG_IV; 1.
 DR Pfam; PF00009; GTP EFTU; 1.
 DR Pfam; PF03144; GTP EFTU D2; 1.
 DR PRINTS; PR00315; ELONGATNFT.
 DR TIGRFAMS; TIGR00484; EF-G; 1.
 DR TIGRFAMS; TIGR00231; small GTP; 1.
 DR PROSITE; PS00301; EFATOR GTP; 1.
 KW Elongation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT NP_BIND 17 24 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SQ SEQUENCE 698 AA; 76927 MW; F44FDA4DB8FC4ECC CRC64;
 Query Match 28.3%; Score 55.5; DB 1; Length 698;
 Best Local Similarity 46.4%; Pred. No. 48;
 Matches 13; Conservative 4; Mismatches 10; Indels 1; Gaps 1;
 QY 14 ANARGD-KSIAIGDIAQAALGSGSIAIGD 40
 DB 366 ANKRDEIKRAGDIAAAGLKDVTGD 393
 RESULT 15
 BDHA_ALCEU STANDARD; PRT; 258 AA.
 ID BDHA_ALCEU
 AC Q9X6U2;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE D-beta-hydroxybutyrate dehydrogenase (EC 1.1.1.30) (BDH)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:33:49 ; Search time 1.32426 Seconds
(without alignments)
2905.511 Million cell updates/sec

Title: US-09-813-214A-11
Perfect score: 196
Sequence: 1 GIGISEADGGKGGANARQKSTAGDTAOLGSOSIAIGD 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs. 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Query			DB	ID	Description
			Match	Length	%			
1	1	67.5	34.4	1588	2	A86036	probable adhesin Z	
2	2	67.5	34.4	1588	2	H9188	probable adhesin E	
3	3	66	33.7	1004	2	C82672	surface-exposed O	
4	4	64.5	32.9	1190	2	A76519	surface protein XF	
5	5	63	32.1	615	2	H70589	hypothetical glyci	
6	6	63	32.1	694	2	F70868	hypothetical glyci	
7	7	61.5	31.4	2342	2	T18200	fatty-acid synthas	
8	8	61	31.1	394	2	B81663	DNA-directed RNA p	
9	9	61	31.1	455	2	S04912	yopA protein - Yer	
10	10	61	31.1	658	2	AOH110	probable surface p	
11	11	60.5	30.9	2059	2	D83671	surface protein XF	
12	12	60	30.6	278	2	F49067	transcription init	
13	13	60	30.6	327	2	T00797	hypothetical prote	
14	14	60	30.6	621	2	T46851	conserved hypothet	
15	15	59.5	30.4	434	2	S04534	invasin precursor	
16	16	58.5	29.8	104	2	T41381	Chaperonins 10 Kd	
17	17	58	29.6	1176	2	T18042	ice nucleation pro	
18	18	57.5	29.3	436	2	T36478	probable solute-bi	
19	19	57.5	29.3	676	1	EDB822	immediate-early pr	
20	20	57.5	29.3	1538	2	H70846	hypothetical glyci	
21	21	57.5	29.3	1660	2	A70869	hypothetical glyci	
22	22	57	29.1	138	2	C87389	hypothetical prote	
23	23	57	29.1	422	2	S04911	yopA protein - Yer	
24	24	57	29.1	505	2	AC3486	cell surface prote	
25	25	57	29.1	1398	2	C87448	hypothetical prote	
26	26	57	29.1	1737	2	A59235	unconventional myo	
27	27	57	29.1	1901	2	F70806	hypothetical glyci	
28	28	56.5	28.8	885	2	F70812	hypothetical glyci	
29	29	56	28.6	362	2	AB3486	cell surface prote	

ALIGNMENTS

RESULT 1
A86036
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A86036
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1588 <STO>
A:Cross-references: GB:AE005174; NID:g12518349; PIDN:AAG58749.1; GSPDB:GN00145; UWGP:Z50
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z5029

30	56	28.6	377	2	A71505	DNA-directed RNA p
31	56	28.6	403	2	A72523	probable acyl-CoA
32	56	28.6	552	2	T23755	hypothetical prote
33	56	28.6	584	2	G70804	hypothetical glyci
34	56	28.6	1107	2	AC0976	probable autotrans
35	56	28.6	1334	2	T50568	probable multi-dom
36	55.5	28.3	698	2	C82332	translation elonga
37	55	28.1	229	2	AB0220	flagellar L-ring p
38	55	28.1	290	2	F87403	hypothetical prote
39	55	28.1	447	2	E83465	conserved hypothet
40	55	28.1	2174	2	E35965	hypothetical glyci
41	54.5	27.8	297	2	AF2605	oxidoreductase Aru
42	54.5	27.8	297	2	E37387	hypothetical prote
43	54.5	27.8	340	2	T20807	hypothetical prote
44	54.5	27.8	644	2	T15652	hypothetical prote
45	54.5	27.8	741	2	G70917	hypothetical glyci

Query Match 32.1%; Score 63; DB 2; Length 694;
 Best Local Similarity 44.4%; Pred. No. 9.7;
 Matches 20; Conservative 3; Mismatches 14; Indels 8; Gaps 3;
 QY 1 GIGISEADGGKGGANARGDKSIAIGDIAQAL-----GSQSIAIG 39
 Db 296 GGGGTAAAGG-GGGNG-GDGGVAQGDIAAFGDCGNGSDGVVAG 338
 RESULT 7
 T18200
 fatty-acid synthase homolog p260 - silkworm
 C:Species: Bombyx mori (silkworm)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
 C:Accession: T18200
 R:Juno, K.
 submitted to the EMBL Data Library, August 1996
 A:Description: P260/270 expressed in embryonic abdominal leg. cells of Bombyx mori can tr
 A:Reference number: Z18815
 A:Accession: T18200
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2342 <UN>
 A:Cross-references: EMBL:U67866; NID:G2058457; PID:G2058458; PIDN:AA853257.1
 C:Superfamily: rat fatty-acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hom
 yrolase homology; short-chain alcohol dehydrogenase homology; [acyl-carrier-protein] S-
 C:Keywords: carrier protein
 Query Match 31.4%; Score 61.5; DB 2; Length 2342;
 Best Local Similarity 40.0%; Pred. No. 48;
 Matches 18; Conservative 5; Mismatches 13; Indels 9; Gaps 2;
 QY 1 GIGISEADG-----GKGANARGDKSIAIGDIAQALGSQSIAIG 39
 Db 322 GAATAEADGKELEAIGKFVAN---DKTIKVGCVKNGSSEAAAG 363
 RESULT 8
 B81663
 DNA-directed RNA polymerase, alpha chain TC0794 [imported] - Chlamydia muridarum (strain
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
 C:Accession: B81663
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: B81500; MUID:20150255; PMID:10684935
 A:Accession: B81663
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-394 <TET>
 A:Cross-references: GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF39597.1; PID:g719082
 A:Experimental source: strain Nigg (MoPn)
 C:Genetics:
 A:Gene: TC0794
 C:Superfamily: DNA-directed RNA polymerase alpha chain
 Query Match 31.1%; Score 61; DB 2; Length 394;
 Best Local Similarity 32.4%; Pred. No. 9.5;
 Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 QY 4 TSEADGGKGGANARGDKSIAIGDIAQALGSQSIAIGD 40
 Db 123 LQDCEGGRASQKLRAITISIDASDLAAAGGQKAITLGD 159
 RESULT 9
 S04912
 yopA protein - Yersinia enterocolitica plasmid pYV6471/76
 C:Species: Yersinia enterocolitica
 C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S04912
 R:Skurnik, M.; Wolf-Watz, H.
 Mol. Microbiol. 3, 517-529, 1989
 A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
 A:Reference number: S04910; MUID:89343638; PMID:2761389
 A:Accession: S04912
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-455 <SKU>
 A:Cross-references: EMBL:X13882; NID:G48606; PIDN:CAA32086.1; PID:G48607
 C:Genetics:
 A:Genome: plasmid pYV6471/76
 Query Match 31.1%; Score 61; DB 2; Length 455;
 Best Local Similarity 46.9%; Pred. No. 11;
 Matches 15; Conservative 6; Mismatches 9; Indels 2; Gaps 1;
 QY 10 GKGG--ANARGDKSIAIGDIAQALGSQSIAIG 39
 Db 59 GAGGLNASAKGIHSIAIGATAEAAKGAAVAG 90
 RESULT 10
 AH0110
 probable surface protein (partial) YPO0902 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C:Accession: AH0110
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, J.
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AH0001; MUID:21470413; PMID:11586360
 A:Accession: AH0110
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-658 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:GL5978974; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO0902
 Query Match 31.1%; Score 61; DB 2; Length 658;
 Best Local Similarity 36.7%; Pred. No. 16;
 Matches 18; Conservative 6; Mismatches 15; Indels 10; Gaps 2;
 QY 1 GIGI-----SEADGGKGG--GANARGDKSIAIGDIAQALGSQSIAIG 39
 Db 167 GVGIGNTALVGAATGGTGAIGTGQTAAAGATAIGSAQAQAQASLALG 215
 RESULT 11
 D82671
 surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82671
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82671
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2059 <SIM>
 A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN001.
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh-Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.U. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XFI529

Query Match 30.9%; Score 60.5; DB 2; Length 2059;
Best Local Similarity 42.5%; Pred. No. 55;
Matches 17; Conservative 3; Mismatches 15; Indels 5; Gaps 1;

QY 5 SEADGGKGGAN-----ARGDKSTAIGDIAQAALGSQSTAIIG 39
Db 1722 STNDGGTGGNDGDGATGSKAIAAGVGTQASGEGAAVG 1761

RESULT 12

A49067
transcription initiation factor IID chain p42 - fruit fly (*Drosophila melanogaster*)
N;Alternate names: TATA-binding protein (TBP)-associated factor TAFII40; TFIID complex C;Species: *Drosophila melanogaster*
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Nov-2000
A;Accession: A49067; S42221
R;Goodrich, J.A.; Hoey, T.; Thut, C.J.; Admon, A.; Tjian, R. Cell 75, 519-530, 1993
A;Title: *Drosophila* TAF-II40 interacts with both a VP16 activation domain and the basal A;Reference number: A49067; MUID:94037099; PMID:8221891
A;Accession: A49067
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-278 <GO>
A;Cross-references: GB:129540; NID:9463048; PID:9463049
R;Kokubo, T.; Gong, D.W.; Wootton, J.C.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y. Nature 367, 484-487, 1994
A;Title: Molecular cloning of *Drosophila* TFIID subunits.
A;Reference number: S42220; MUID:94150630; PMID:7545910
A;Accession: S42221
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-278 <XOK>
A;Cross-references: EMBL:U06458; NID:9458679; PIDN:AAC47347.1; PID:9458680
A;Gene: FlyBase:Taf40
A;Cross-references: FlyBase:FBgn0000617
C;Keywords: transcription initiation

Query Match 30.6%; Score 60; DB 2; Length 278;
Best Local Similarity 31.8%; Pred. No. 8.9;
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 2 IGISEADGGKGGANARGDKSTAIGDIAQAALGSQSTAIIG 39
Db 198 VGSSGGSGGGGGQGVKSESTGAGDLKWEVDSDAAAVG 235

RESULT 13

T00797
hypothetical protein At2g32710 [imported] - *Arabidopsis thaliana*
N;Alternate names: hypothetical protein F24L7.15
C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
A;Accession: T00797; E84736
R;Rounsley, S.D.; Lin, X.D.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul submitted to the EMBL Data Library, February 1998
A;Description: *Arabidopsis thaliana* chromosome II BAC F24L7 genomic sequence.
A;Reference number: Z14204
A;Accession: T00797

A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-327 <ROU>
A;Cross-references: EMBL:AC003974; NID:92914688; PID:92914702
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84736
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-327 <STO>
A;Cross-references: GB:AE002093; NID:92914702; PIDN:AAC04492.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32710; F24L7.15
A;Map position: 2
A;Introns: 193/2
C;Superfamily: *Arabidopsis thaliana* hypothetical protein F24L7.15

Query Match 30.8%; Score 60; DB 2; Length 327;
Best Local Similarity 33.3%; Pred. No. 10;
Matches 12; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY 3 GISEADGGKGGANARGDKSTAIGDIAQAALGSQSTAIIG 38
Db 12 GAGAGAGGGGGGGGESSIALMDVWVSPSSSLGV 47

RESULT 14

T46851
conserved hypothetical protein ybaU [imported] - *Rhodobacter sphaeroides*
C;Species: *Rhodobacter sphaeroides*
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
A;Accession: T46851
R;Mackenzie, C.; Simmons, A.E.; Kaplan, S. Genetics 153, 525-538, 1999
A;Title: Multiple chromosomes in bacteria. The Yin and Yang of trp gene localization in A;Reference number: Z24108; MUID:99442363; PMID:10511537
A;Accession: T46851
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-621 <MAC>
A;Cross-references: EMBL:AF108766; NID:94185542; PIDN:AAD09115.1; PID:94185543
A;Experimental source: strain 2.4.1
C;Genetics:
A;Gene: ybaU

Query Match 30.6%; Score 60; DB 2; Length 621;
Best Local Similarity 33.3%; Pred. No. 20;
Matches 13; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 1 GIGISPADGGKGGANARGDKSTAIGDIAQAALGSQSTAIIG 39
Db 34 GFGVTFGGGLSTVGSGDRDITTGAVARALQEMRALG 72

RESULT 15

S04534
invasin precursor - *Yersinia pseudotuberculosis* plasmid pIBI
C;Species: *Yersinia pseudotuberculosis*
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
A;Accession: S04534; S04910
R;Rosqvist, R.; Skurnik, M.; Wolf-Watz, H. Nature 334, 522-525, 1998
A;Title: Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations A;Reference number: S04534; MUID:88302441; PMID:3043229
A;Accession: S04534
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <ROS>

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:24:53 ; Search time 4.11791 Seconds
(without alignments)
2744.569 Million cell updates/sec

Title: US-09-813-214A-11

Perfect score: 196

Sequence: 1 GIGISEADGGKGNARGDKSIAIGDIAQLGQSIAIGD 40

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	196	100.0	40	4	AAE00705	-----	Aae00705 N-termina
2	196	100.0	2122	6	ABU08784		Abu08784 Moraxella
3	196	100.0	2123	4	AAE00701		Aae00701 Moraxella
4	190	96.9	43	2	AAW32274		Aaw32274 M. catarr
5	190	96.9	43	4	AAE00703		Aae00703 N-termina
6	121	61.7	24	2	AAW32276		Aaw32276 M. catarr
7	121	61.7	24	4	AAE00702		Aae00702 Internal
8	86.5	44.1	1946	6	ABU35023		Abu35023 Protein e
9	84	42.9	1992	2	AAW04505		Aaw04505 Moraxella
10	84	42.9	1992	4	AAE69137		Aae69137 M. catarr
11	84	42.9	1992	4	AAE69133		Aae69133 M. catarr
12	84	42.9	2047	4	AAE69134		Aae69134 M. catarr
13	84	42.9	2053	4	AAE69135		Aae69135 M. catarr
14	84	42.9	2139	6	ABP71294		Abp71294 M. catarr
15	84	42.9	2314	4	AAE69136		Aae69136 M. catarr
16	75	38.3	2265	6	ABU17199		Abu17199 Protein e
17	75	38.3	2504	6	ADA34534		Ada34534 Acinetoba
18	74	37.8	3073	6	ABU21223		Abu21223 Protein e
19	67.5	34.4	1588	7	ADC01413		Adc01413 Enterohae
20	67	34.2	1778	4	ABE52677		AbE52677 Escherich
21	67	34.2	2042	2	AAW56319		Aaw56319 Haemophil
22	66	33.7	261	2	AAI13486		Aai13486 Mouse clo
23	66	33.7	435	6	ABP76894		Abp76894 N. gonorr
24	66	33.7	2712	6	ABU39146		Abu39146 Protein e
25	65.5	33.4	270	6	ABU19868		Abu19868 Protein e

ALIGNMENTS

RESULT 1

AAE00705

ID AAE00705 standard; peptide; 40 AA.

XX AAE00705;

XX 02-JUL-2001 (first entry)

XX N-terminal #2 of M. catarrhalis outer membrane protein-106 (OMP106).

XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;

XX bacterial infection; immunogen; cytotoxic; antibiotic;

XX passive immunisation.

XX Moraxella catarrhalis.

XX US6214981-B1.

XX 10-APR-2001.

XX 12-NOV-1997; 97US-00968685.

XX 03-MAY-1996; 96US-00642712.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L, Tillman UP;

XX WPI; 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane protein-106 polypeptide, useful for diagnosis of bacterial infections and as vaccine against Moraxella catarrhalis infection of mammals.

XX Example; Col 31; 49pp; English.

XX The present sequence is N-terminal of haemagglutinating Moraxella catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a therapeutic and prophylactic vaccine against M. catarrhalis infections of mammals. It is used for diagnosis of bacterial infections and as reagents for clinical or medical diagnosis of M. catarrhalis infections and for scientific research on the properties of pathogenicity, virulence and infectivity of M. catarrhalis. It is also used as a probe to identify the presence of M. catarrhalis in biological specimens and to identify other bacteria that encode a polypeptide related to M. catarrhalis OMP106. OMP106-derived polypeptides are used as ligands to detect antibodies elicited in response to M. catarrhalis infections and also as immunogens for inducing M. catarrhalis-specific antibodies which are useful in

26	65	33.2	1461	6	ABU47415	Protein e
27	64	32.7	1099	6	ABU20296	Protein e
28	64	32.7	1400	6	ABU45394	Protein e
29	63	32.1	505	6	ABU34417	Protein e
30	63	32.1	615	6	ABU36862	Protein e
31	63	32.1	694	6	ABU36802	Protein e
32	63	32.1	1129	6	ABU20124	Protein e
33	61	31.1	394	6	ABU27363	Protein e
34	61	31.1	2039	2	AAW56322	Haemophil
35	60	30.6	278	2	AAW56490	TATA-bind
36	60	30.6	278	2	AAW06081	Drosophil
37	60	30.6	278	2	AAW25025	TATA-bind
38	60	30.6	278	4	ABB62342	Arabidops
39	60	30.6	289	3	AAE27262	Drosophil
40	60	30.6	289	3	AAE27262	Arabidops
41	60	30.6	289	5	AAE25109	Arabidops
42	60	30.6	289	5	AAU72581	Arabidops
43	60	30.6	327	3	AAE27261	Arabidops
44	60	30.6	449	4	ABB66828	Drosophil
45	59	30.1	462	6	ABR57117	ScncoPFCA

CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis
 XX
 SQ Sequence 40 AA;

Query Match 100.0%; Score 196; DB 4; Length 40;
 Best Local Similarity 100.0%; Pred. No. 6.7e-18;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIGISADGKGANARGDKSIAIGDIAQALGQSSTAIGD 40
 |||||
 Db 1 GIGISADGKGANARGDKSIAIGDIAQALGQSSTAIGD 40

RESULT 2

ABU08784
 ID ABU08784 standard; protein; 2122 AA.

AC ABU08784;

XX 28-MAY-2003 (first entry)

DE Moraxella catarrhalis outer membrane protein, OMP106.

XX Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;
 KW vaccine.

OS Moraxella catarrhalis.

XX Key Location/Qualifiers

FT Peptide 1..68

FT /label= signal_sequence

FT Protein 69..2122

FT /label= Mature_OMP106

FT Region 69..111

FT /label= N-terminal sequence of mature OMP106

FT /note= "Specifically claimed in claim 9"

FT Misc-difference 779..880

FT /note= "Encoded by ATCTAAGGC"

XX US2002177200-A1.

XX 28-NOV-2002.

XX 20-MAR-2001; 2001US-00813214.

XX 03-MAY-1996; 96US-00642712.

XX 12-NOV-1997; 97US-00968685.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L;

XX WPI; 2003-328486/31.

XX N-PSDB; ABX93525.

XX Novel outer membrane polypeptide of Moraxella catarrhalis, useful for
 PT producing an immune response in an animal, and as ligands to detect
 PT antibodies elicited in response to Moraxella infections.

XX Claim 33; Page 28-34; 5lpp; English.

XX The invention relates to an isolated or substantially pure outer membrane
 CC polypeptide of Moraxella catarrhalis (OMP106 polypeptide). The
 CC polypeptide and its peptide fragments are useful for producing an immune
 CC response in an animal, and as ligands to detect antibodies elicited in
 CC response to M. catarrhalis infections. The polypeptide and its peptide
 CC fragment are further useful as active ingredients in vaccines against M.
 CC catarrhalis infections. The polypeptide is useful to prepare antibodies.
 CC An antibody binding the polypeptide is useful in immunoassays to detect
 CC M. catarrhalis in biological specimens, and also in passive immunisations
 CC against M. catarrhalis infections. An antibody binding the polypeptide is
 CC also useful to facilitate isolation and purification of the polypeptide

CC and its peptide fragment, and as probes for identifying clones in
 CC expression libraries that have inserts encoding the polypeptide and
 CC peptide fragment. An antibody binding the polypeptide is also useful to
 CC diagnose M. catarrhalis infections. The polynucleotide encoding the
 CC polypeptide is useful as a probe to identify the presence of M.
 CC catarrhalis in biological specimens by hybridisation or PCR amplification
 CC and also to detect other bacteria that might encode a polypeptide related
 CC to M. catarrhalis OMP106. The polypeptide, an antibody binding the
 CC polypeptide, its peptide fragment and the polynucleotide encoding the
 CC polypeptide are useful as reagents for clinical medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis, as well as
 CC host defence mechanisms. The present sequence represents the amino acid
 CC sequence of M. catarrhalis outer membrane protein (OMP)-106

XX Sequence 2122 AA;

Query Match 100.0%; Score 196; DB 6; Length 2122;

Best Local Similarity 100.0%; Pred. No. 5.7e-16;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIGISADGKGANARGDKSIAIGDIAQALGQSSTAIGD 40

|||||
 Db 68 GIGISADGKGANARGDKSIAIGDIAQALGQSSTAIGD 107

RESULT 3

AAE00701

ID AAE00701 standard; protein; 2123 AA.

XX AAE00701;

XX 02-JUL-2001 (first entry)

XX Moraxella catarrhalis outer membrane protein-106 (OMP106).

XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.

XX Moraxella catarrhalis.

XX US6214981-B1.

XX 10-APR-2001.

XX 12-NOV-1997; 97US-00968685.

XX 03-MAY-1996; 96US-00642712.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L, Tillman UF;

XX WPI; 2001-281002/29.

XX N-PSDB; AAD04029.

XX Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.

XX Claim 7; Col 53-64; 49pp; English.

XX The present sequence is haemagglutinating Moraxella catarrhalis outer
 CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic and
 CC prophylactic vaccine against M. catarrhalis infections of mammals. It is
 CC used for diagnosis of bacterial infections and as reagents for clinical
 CC or medical diagnosis of M. catarrhalis infections and for scientific
 CC research on the properties of pathogenicity, virulence and infectivity of
 CC M. catarrhalis. It is also used as a probe to identify the presence of M.
 CC catarrhalis in biological specimens and to identify other bacteria that
 CC encode a polypeptide related to M. catarrhalis OMP106. OMP106-derived
 CC polypeptides are used as ligands to detect antibodies elicited in

CC response to M. catarrhalis infections and also as immunogens for inducing
 CC M. catarrhalis-specific antibodies which are useful in immunoassays to
 CC detect M. catarrhalis in biological specimens. Cytotoxic antibodies are
 CC useful in passive immunisations against M. catarrhalis

XX SQ Sequence 2123 AA;

Query Match 100.0%; Score 196; DB 4; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 5.7e-16; Mismatches 0; Indels 0; Gaps 0;
 Matches 40; Conservative 0;

QY 1 GIGISEADGGKGGANARGDKSIAIGDIAQALGQSQIAIGD 40
 |||||
 Db 68 GIGISEADGGKGGANARGDKSIAIGDIAQALGQSQIAIGD 107

RESULT 4

AAW32274
 ID AAW32274 standard; peptide; 43 AA.

XX AC AAW32274;

XX DT 08-MAY-1998 (first entry)

XX DE M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.

XX KW Outer membrane protein-106; OMP106; vaccine; immune response;
 XX KW cytotoxic antibody; Moraxella catarrhalis.

XX OS Moraxella catarrhalis.

XX PN WO9741731-A1.

XX PD 13-NOV-1997.

XX PF 28-APR-1997; 97WO-US007679.

XX PR 03-MAY-1996; 96US-00642712.

XX PA (ANTE-) ANTEX BIOLOGICS INC.

XX PI Tucker K, Plosila L;

XX DR WPI; 1997-558601/51.

XX PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 XX PT vaccines for producing immune responses against M. catarrhalis.

XX PS Claim 9; Page 23; 78pp; English.

XX CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella
 CC catarrhalis, an haemagglutinating cultivar. The peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of M.
 CC catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis

XX SQ Sequence 43 AA;

Query Match 96.9%; Score 190; DB 2; Length 43;
 Best Local Similarity 100.0%; Pred. No. 4.2e-17; Mismatches 0; Indels 0; Gaps 0;
 Matches 39; Conservative 0;

QY 2 IGISEADGGKGGANARGDKSIAIGDIAQALGQSQIAIGD 40
 |||||
 Db 1 IGISEADGGKGGANARGDKSIAIGDIAQALGQSQIAIGD 39

RESULT 5

AAE00703

ID AAE00703 standard; peptide; 43 AA.

XX AC AAE00703;

XX DT 02-JUL-2001 (first entry)

XX DE N-terminal #1 of M. catarrhalis outer membrane protein-106 (OMP106).

XX KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 XX KW bacterial infection; immunogen; cytotoxic; antibiotic;
 XX KW passive immunisation.

XX OS Moraxella catarrhalis.

XX PN US6214981-B1.

XX PD 10-APR-2001.

XX PF 12-NOV-1997; 97US-00968685.

XX PR 03-MAY-1996; 96US-00642712.

XX PA (ANTE-) ANTEX BIOLOGICS INC.

XX PI Tucker K, Plosila L, Tillman UF;

XX DR WPI; 2001-281002/29.

XX PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 XX PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 XX PT as vaccine against Moraxella catarrhalis infection of mammals.

XX PS Claim 2; Col 31; 49pp; English.

XX CC The present sequence is N-terminal of haemagglutinating Moraxella
 CC catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a
 CC therapeutic and prophylactic vaccine against M. catarrhalis infections of
 CC mammals. It is used for diagnosis of bacterial infections and as reagents
 CC for clinical or medical diagnosis of M. catarrhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarrhalis. It is also used as a probe to identify the
 CC presence of M. catarrhalis in biological specimens and to identify other
 CC bacteria that encode a polypeptide related to M. catarrhalis OMP106.

XX CC OMP106-derived polypeptides are used as ligands to detect antibodies
 CC elicited in response to M. catarrhalis infections and also as immunogens
 CC for inducing M. catarrhalis-specific antibodies which are useful in
 CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis

XX SQ Sequence 43 AA;

Query Match 96.9%; Score 190; DB 4; Length 43;
 Best Local Similarity 100.0%; Pred. No. 4.2e-17; Mismatches 0; Indels 0; Gaps 0;
 Matches 39; Conservative 0;

QY 2 IGISEADGGKGGANARGDKSIAIGDIAQALGQSQIAIGD 40
 |||||
 Db 1 IGISEADGGKGGANARGDKSIAIGDIAQALGQSQIAIGD 39

RESULT 6

AAW32276

ID AAW32276 standard; peptide; 24 AA.

XX AC AAW32276;

XX DT 08-MAY-1998 (first entry)

XX DE M. catarrhalis outer membrane protein (OMP)-106 partial peptide fragment.

XX KW Outer membrane protein-106; OMP106; vaccine; immune response;
 XX KW cytotoxic antibody; Moraxella catarrhalis; primer; probe.

XX OS Moraxella catarrhalis.

XX WO9741731-A1.
 PN
 XX
 PD 13-NOV-1997.
 XX
 XX 28-APR-1997; 97WO-US007679.
 XX
 XX 03-MAY-1996; 96US-00642712.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 XX Tucker K, Flosila L;
 PI
 XX WPI; 1997-558601/51.
 XX
 DR N-PSDB; AAT86522.
 XX
 XX Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 PT vaccines for producing immune responses against M. catarrhalis.
 PT
 XX Disclosure; Page 58; 78pp; English.
 PS
 XX This is a partial sequence of an amino terminal peptide fragment of a
 CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an
 CC outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating
 CC cultivar. The encoding DNA can be used as a 5' primer for PCR
 CC amplification of a full length OMP106 DNA. The DNA fragment can also be
 CC used as a probe for screening M. catarrhalis genomic libraries for OMP106
 CC polypeptide coding sequences. The encoded peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of M.
 CC catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis
 CC
 XX Sequence 24 AA;
 SQ

Query Match 61.7%; Score 121; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 EADGGKGGANARGDKSIAIGDIAQ 29
 Db 1 EADGGKGGANARGDKSIAIGDIAQ 24
 |||||

RESULT 7
 AAE00702
 ID AAE00702 standard; peptide; 24 AA.
 XX
 XX AAE00702;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE Internal fragment of N-terminal outer membrane protein-106 (OMP106).
 XX
 KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.
 XX
 XX US6214981-B1.
 PN
 XX
 XX 10-APR-2001.
 PD
 XX
 PF 12-NOV-1997; 97US-00968685.
 XX
 XX 03-MAY-1996; 96US-00642712.
 XX
 XX (ANTE-) ANTEX BIOLOGICS INC.
 PA
 XX Tucker K, Flosila L, Tillman UP;
 PI
 XX

DR WPI; 2001-281002/29.
 DR N-PSDB; AAD04030.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 XX Example; Col 43-44; 49pp; English.
 PS
 XX
 CC The present sequence is the internal fragment of N-terminal
 CC haemagglutinating Moraxella catarrhalis outer membrane protein-106
 CC (OMP106). This sequence is used to design a probe and a 5' PCR primer.
 CC The OMP106 is used as a therapeutic and prophylactic vaccine against M.
 CC catarrhalis infections of mammals. It is used for diagnosis of bacterial
 CC infections and as reagents for clinical or medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also
 CC used as a probe to identify the presence of M. catarrhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used as
 CC ligands to detect antibodies elicited in response to M. catarrhalis
 CC infections and also as immunogens for inducing M. catarrhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarrhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarrhalis
 CC
 XX Sequence 24 AA;
 SQ

Query Match 61.7%; Score 121; DB 4; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 EADGGKGGANARGDKSIAIGDIAQ 29
 Db 1 EADGGKGGANARGDKSIAIGDIAQ 24
 |||||

RESULT 8
 ABU35023
 ID ABU35023 standard; protein; 1946 AA.
 XX
 XX ABU35023;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #20550.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Moraxella catarrhalis.
 XX
 XX WO200277183-A2.
 PN
 XX
 PD 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 PR
 PR 06-SEP-2001; 2001US-00948993.
 PR
 PR 23-OCT-2001; 2001US-0342923P.
 PR
 PR 08-FEB-2002; 2002US-00072851.
 PR
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HJ;
 PI
 XX WPI; 2003-029926/02.
 DR
 DR N-PSDB; ACA38893.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 XX

CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the M.
 CC catarrhalis M56 200kDa protein in pKS348, which is given in the
 CC exemplification of the present invention
 XX
 XX SQ Sequence 1992 AA;

Query Match 42.9%; Score 84; DB 4; Length 1992;

Best Local Similarity 56.4%; Pred. No. 0.11; Indels 4; Gaps 1;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR---GDKSIAIGDIAQAALGSGSIAIG 39

Db 1718 AKADGEAAVAIGRTQAGNQSIAIGDVAQATGDSIAIG 1756

RESULT 11

AAB69133

ID AAB69133 standard; protein; 1992 AA.

AC AAB69133;

XX 24-APR-2001 (first entry)

DT M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

DE Moraxella catarrhalis strain 4223; major outer membrane protein;

XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection.

KW Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA000870.

XX 27-JUL-1999; 99US-00361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59100, AAF59101.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis.

XX Example 3; Fig 2A-W; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

XX The 200 kDa outer membrane protein (II) has antibacterial activity and

XX can be used in vaccines. (II), and its truncated versions, are used as

XX immunogenic compositions and vaccines to protect against M. catarrhalis

XX infections, particularly otitis media in humans. (II) is also used as

XX antigen in immunoassays for detecting specific antibodies (Ab), and to

XX generate Ab. (I) are used for recombinant production of (II) and its

XX fragments are used as probes for identifying/cloning 200 kDa protein

XX genes from other strains, and for diagnostic detection of M. catarrhalis.

XX (I) makes possible production of large amount of recombinant immunogens.

XX Expression of truncated versions of (II) reduces toxicity of the protein

XX towards the Escherichia coli host. The present sequence represents the M.

XX catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is used

CC in the exemplification of the present invention

XX SQ Sequence 1992 AA;

Query Match 42.9%; Score 84; DB 4; Length 1992;

Best Local Similarity 56.4%; Pred. No. 0.11;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR---GDKSIAIGDIAQAALGSGSIAIG 39

Db 1718 AKADGEAAVAIGRTQAGNQSIAIGDVAQATGDSIAIG 1756

RESULT 12

AAB69134

ID AAB69134 standard; protein; 2047 AA.

XX AAB69134;

XX 24-APR-2001 (first entry)

DT M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA000870.

XX 27-JUL-1999; 99US-00361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59102, AAF59103.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis.

XX Claim 1; Fig 3A-W; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

XX The 200 kDa outer membrane protein (II) has antibacterial activity and

XX can be used in vaccines. (II), and its truncated versions, are used as

XX immunogenic compositions and vaccines to protect against M. catarrhalis

XX infections, particularly otitis media in humans. (II) is also used as

XX antigen in immunoassays for detecting specific antibodies (Ab), and to

XX generate Ab. (I) are used for recombinant production of (II) and its

XX fragments are used as probes for identifying/cloning 200 kDa protein

XX genes from other strains, and for diagnostic detection of M. catarrhalis.

XX (I) makes possible production of large amount of recombinant immunogens.

XX Expression of truncated versions of (II) reduces toxicity of the protein

XX towards the Escherichia coli host. The present sequence represents the M.

XX catarrhalis strain 4223 genomic 200kDa protein, which is given in the

XX exemplification of the present invention

XX SQ Sequence 2047 AA;

Query Match 42.9%; Score 84; DB 4; Length 2047;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR---GDKSIAIGDIAQAALGSGSIAIG 39

Db 1718 AKADGEAAVAIGRTQAGNQSIAIGDVAQATGDSIAIG 1756

Db 1773 AKADGEAAVAIGRTQAGNQSTAIAGDQATGDSIAIG 1811

RESULT 13

AAB69135
ID AAB69135 standard; protein; 2053 AA.

XX AC AAB69135;

XX DT 24-APR-2001 (first entry)

XX DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.

XX KW Moraxella catarrhalis strain Q8; major outer membrane protein;

XX KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX KW otitis media; detection.

XX OS Moraxella catarrhalis.

XX PN WO200107619-A1.

XX PD 01-FEB-2001.

XX PF 26-JUL-2000; 2000WO-CA000870.

XX PR 27-JUL-1999; 99US-00361619.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX DR WPI; 2001-159722/16.

XX DR N-PSDB; AAF59104.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis.

XX Claim 1; Fig 4A-V; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the M.
CC catarrhalis strain Q8 200kDa protein, which is given in the
XX exemplification of the present invention

XX SQ Sequence 2053 AA;

Query Match 42.9%; Score 84; DB 4; Length 2053;
Best Local Similarity 56.4%; Pred. No. 0.12;
Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR----GDKSTAIAGDIAQALGSQSTAIAG 39

Db 1779 AKADGEAAVAIGRTQAGNQSTAIAGDQATGDSIAIG 1817

RESULT 14

ABP71294
ID ABP71294 standard; protein; 2139 AA.

XX AC ABP71294;

XX DT 28-APR-2003 (first entry)

XX DE M. catarrhalis surface exposed IgD binding protein.

XX KW Surface exposed protein; bacterium; immunoglobulin; IgD; immunomodulator;
XX gene therapy; vaccine; mid.

XX OS Moraxella catarrhalis.

XX FH Key Location/Qualifiers

FT Peptide 1..66

FT Protein /note= "signal peptide"

FT Protein 67..2139

FT Region /note= "mature protein"

FT Region 764..913

FT Region /note= "specifically claimed immunogenic fragment"

FT Region 962..1200

FT Region /note= "specifically claimed immunogenic fragment"

XX PN WO2003004651-A1.

XX PD 16-JAN-2003.

XX PF 01-JUL-2002; 2002WO-SE001299.

XX PR 04-JUL-2001; 2001SE-00002410.

XX PA (FORS/) FORSGREN A.

XX XX

XX PI Forsgren A, Riesbeck K, Janson H;

XX DR WPI; 2003-221598/21.

XX DR N-PSDB; ABZ58968.

PT New surface exposed immunoglobulin D-binding protein from Moraxella

PT catarrhalis, useful for treating an autoimmune disease or as vaccine,

PT comprises a molecular weight of 200 kDa.

XX Claim 1; Fig 6; 98pp; English.

CC The invention relates to a surface exposed immunoglobulin D-binding
CC protein detected in Moraxella catarrhalis. The protein has an apparent
CC molecular weight of 200 kDa, and can selectively bind membrane bound or
CC soluble IgD. The protein or its variant or fragment, is useful in
CC treating an autoimmune disease or as vaccine. The present sequence
CC represents a M. catarrhalis the surface exposed protein (mid) of the
CC invention

XX SQ Sequence 2139 AA;

Query Match 42.9%; Score 84; DB 6; Length 2139;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR----GDKSTAIAGDIAQALGSQSTAIAG 39

Db 1865 AKADGEAAVAIGRTQAGNQSTAIAGDQATGDSIAIG 1903

RESULT 15

AAB69136

ID AAB69136 standard; protein; 2314 AA.

XX AC AAB69136;

XX DT 24-APR-2001 (first entry)

XX DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.

XX KW Moraxella catarrhalis strain Q8; major outer membrane protein;

XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX KW otitis media; detection.

XX OS Moraxella catarrhalis.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:36:00 ; Search time 215.568 Seconds
(without alignments)
3105.882 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 MNHIYKVFINKATGTFMAVA.....NGSADTQGHVGAAGAGGPHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7316	68.3	1964	2	Q8QM9	Q8qm9 moraxella c
2	6475	60.5	2314	2	Q8QM8	Q8qm8 moraxella c
3	1246.5	11.6	2059	16	Q9PD50	Q9pd50 xylella fas
4	1233	11.5	2052	16	Q87D81	Q87del xylella fas
5	1185.5	11.1	2353	2	P71401	P71401 haemophilus
6	1063	9.9	1778	16	Q8FCB2	Q8fcb2 escherichia
7	1039	9.7	2712	16	Q9F3X5	Q9f3x5 pasteurella
8	886	9.2	1588	16	Q8XDG4	Q8xdg4 escherichia
9	945	8.8	1004	2	Q8QM77	Q8qm77 haemophilus
10	939	8.8	1002	2	Q8QM78	Q8qm78 haemophilus
11	919.5	8.6	1461	16	Q8ZL64	Q8z164 salmonella
12	847.5	7.9	1315	16	Q87D62	Q87d62 xylella fas
13	831	7.8	1953	16	Q8BH22	Q8bhj2 rhizobium l
14	784.5	7.3	2351	16	Q8PC05	Q8pcq5 xanthomonas
15	763	7.1	1210	2	Q8GM75	Q8gm75 haemophilus
16	763	7.1	2190	16	Q8PGS0	Q8pgs0 xanthomonas

17	761	7.1	1204	2	Q8GM76	Q8gm76 haemophilus
18	756	7.1	1210	2	Q8GM74	Q8gm74 haemophilus
19	730.5	6.8	1299	16	Q9F3X6	Q9f3x6 pasteurella
20	728	6.8	1190	16	Q9PC04	Q9pc04 xylella fas
21	722	6.7	3930	16	Q98E20	Q98e20 rhizobium l
22	684	6.4	3420	16	Q8FUS1	Q8fus1 brucella su
23	682	6.4	1098	2	Q48152	Q48152 haemophilus
24	675.5	6.3	3692	2	Q8KRR3	Q8krr3 fusobacteri
25	670	6.3	1096	2	Q8GM79	Q8gm79 haemophilus
26	670	6.3	1107	16	Q9F2D8	Q9f2d8 salmonella
27	670	6.3	1116	16	Q83ST9	Q83st9 salmonella
28	669	6.2	2201	2	Q8GF46	Q8gf46 zymomonas m
29	669	6.2	3705	16	Q8ZHA1	Q8zhal yersinia pe
30	669	6.2	3710	16	Q8CZU2	Q8czu2 yersinia pe
31	658.5	6.1	3165	16	Q8RDQ9	Q8rdq9 fusobacteri
32	654.5	6.1	3705	2	Q9F285	Q9f285 yersinia pe
33	653	6.1	3133	2	Q8KER1	Q8krr1 fusobacteri
34	651.5	6.1	2340	16	Q9ZD91	Q9zd91 rickettsia
35	650.5	6.1	6274	16	Q880E1	Q880el pseudomonas
36	646	6.0	1264	2	Q8RQ61	Q8rg61 actinobacill
37	645.5	6.0	5291	16	Q8X2T1	Q8x2t1 escherichia
38	643.5	6.0	2411	16	Q8PF72	Q8pf72 xanthomonas
39	636.5	5.9	3961	16	Q8BP94	Q8bp94 xanthomonas
40	635.5	5.9	990	16	Q83J22	Q83j22 shigella fl
41	634	5.9	5188	16	Q8X4H5	Q8x4h5 escherichia
42	630.5	5.9	3322	16	Q8XQZ5	Q8xqz5 raistonia s
43	628	5.9	3441	16	Q89PB9	Q89pb9 bradyrhizob
44	624	5.8	1309	16	Q8XPL8	Q8xpl8 raistonia s
45	623.5	5.8	2806	16	Q8RI19	Q8ril9 fusobacteri

ALIGNMENTS

RESULT 1

Q8QM9 PRELIMINARY; PRT; 1964 AA.
ID Q8QM9
AC Q8QM9
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAG
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O35E;
RX MEDLINE=22112901; PubMed=12117964;
RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Gene J.W. III,
RA Hansen E.J.;
RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
RT Activities.";
RL Infect. Immun. 70:4523-4533 (2002).
DR EMBL; AY077637; AAL78284.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; YADA.
DR Pfam; PF05658; Hep Hag; 9.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 1964 AA; 201585 MW; 764A08F5F1F6854E CRC64;

Query Match

Best Local Similarity 68.3%; Score 7316; DB 2; Length 1964;

Matches 1541; Conservative 127; Mismatches 209; Indels 332; Gaps 29;

QY 1 MNHIYKVFINKATGTFMAVAEYAKSHSTGGSCATGQGVSVRTLSFARIALAVLIGAT 60

DB 1 MNHIYKVFINKATGTFMAVAEYAKSHSTGGSCATGQGVSVRTLSFARIALAVLIGAT 59

Qy	380	TIT-----GGAQTSA	LTHNIGVVQNGD-----GLKVQLAET-----LTSLKMYTTE	421
Dd	336	TITKAI	GSNDVQFNLNKDVKTITLAVGDALLNTDGIALGTDSLSLTCLAITDGPVATAS	395
Qy	422	NLTANEKV-----TVGKTR	LTLTKTG-----FTDMNGIDES-----	453
Dd	396	GIDAGSKVISHVAAGA	VSET--STDVANGSOLNAVQVASQPVTFTGNEGAVKSLGQSV	453
Qy	454	-----KPYLDKDTG-IHA-----GGOKITKLTA	VGVDDD	481
Dd	454	VISGESSTAGTYS	SGNLKSVDDEAAGRILHLQADSPKFGNVVINNGKISGVITAGTEED	513
Qy	482	AATYGOLKKVNQTA	BESALOTFTVKVKDKNGNDANDSKLIITVGONKPEDPGTQVNTLKLKGE	541
Dd	514	AVNFSQLKSISTAVD---QGWT	LITASGNG---SKVASGG-----TVDLKNT	554
Qy	542	NGVDVTTETNG- -TVTFGL	NON- -NGLTVGNSFLANDGLSVKNTSNKQIOVGADGIT	595
Dd	555	DG-NLTTISKSDNDVFNFSK	FQDGMTSGITTVANNQVKV-----GSDVALGITGTLT	608
Qy	596	FTDISNKPAGIEN	TIIRTDGIGIFANNTGSLDANKPRLTPTGINAGGKELTNVQSAL-	654
Dd	609	ITD-----	-GPAVTASGIDAGSKVISHVAAGV	635
Qy	655	NPATNGQOLFPMRLSTANT	EKGSAATIKDLYNLSOV-----PLTFAGDTGPNVTK	706
Dd	636	SETSTDVANGSQL-----NAVQ	QASQPVTFGTGNEGA-VKR	670
Qy	707	KLGRI	LKVGKTKTADDLTKNNIGVADVSTDNSLTVKLAKT-----	747
Dd	671	SLGOSVVISGESSTAGMYS	GGNLKSVDDEAAGRILHLQADSPKFGNVVINNGKISGVTA	730
Qy	748	LSLDLDAVNTKTL-----TASDK- -VTVDSGNNTAKL-----ONGDLTTS	KQNTGA	790
Dd	731	GTEETDAVNFSQLKSISTAVDOGWTLTASGNSGKVASGGTVDLNKDGNL	LTISKSDSN	790
Qy	791	TPAPN-SKITIGVDGLKFTPDNN	IADGTTVIYTKD--KVGFAKQDGLSKPKPLYDRDKLK	847
Dd	791	DWVPNL	SKDFKVDGM-----TSGTTVVNNDGVKVG-----SDVALGTTGTLT	831
Qy	848	VGE-VELTTINGINAGKA	ITG-----LSNLTLDATNATTCHVTQLGIIVDSTDKTRAASIG	901
Dd	832	ITDGPVATASGIDAGSKVISHVAAGV	VSETSTDAVNG-----SOLNAVQASQPVFTFG	886
Qy	902	DVLNAGFNLKNNGDAKFVS	TYDTVDPIVINGNATTAKVTYDGKASKVAYDVNVDGTTIHLT	961
Dd	887	--NEGAVKSLGQSV-----VISSES	TAG-TYSGNLKSVDDEAAG-RHLLQ	930
Qy	962	GADGNKNIQVKTITLT	KTDAKDAINFVNSGDDKALINAKDIADNLNLAGEIRN-T	1020
Dd	931	LADSPK-FG-----NVVNING-----GKISGVT	952	
Qy	1021	KGTADT-ALQTFQVKVKENGDDDDNDADTTV	GKADKTNQNTLKLKXGNGLDIQTNKOG	1079
Dd	953	AGTEETDAVNFSQLKSIT-----STA	VDQGWTLTASGANSKVASG--G	993
Qy	1080	TVTFIGINTQSLKAGNNTILNN	GLSITKNTAGNEQIOVGADG-KVEKAYNNGVVGAGID	1137
Dd	994	TV-----DUKNTDGNL	LTISKSDNDVFPNLKDFKVGMTS	1030
Qy	1138	GTTIRITRDEICFA-----GTNGSLDKSPHL	SKDGINAGGKITTINOSSGIAONSMDAVT	1192
Dd	1031	GTTVVNNDGVKVGSDVALGTTGLTIANGP	AVTASGIDAGSKVISHVAAGVSETSDAVN	1090
Qy	1193	GGKYIDLKTELKENKIS	TAKQNLSHFVSVADEQQNNFTVSNPYSSTYKTS	1252
Dd	1091	GSOLNAVQVASQPVTFGT-----NEGAVK	SLGQSVVIS-----GESSTAGTYS	1135
Qy	1253	GENGIITTKUNKGVVRV	GIDQKTLTTPKL--TVGNNGKGIIVDSQNGQNTITGLSNTLA	1310
Dd	1136	GGN-LKSV	DEAAGRIHLQALA-----DSPKFGNVINNGGK-----ISGVT----	1175

Qy	1311	NVTNDKSVRTTEGKIIKDEBCKTRAAISIVDVLSAGFNLOGNEAVDFVSTYDTVNF--A	1366
Db	1176	-----AGTETDAV--NFSQLKSISTAVDQGGNTLTASGANGSKVASCGGTVDLKNT	1223
Qy	1369	DGNATTAKTVYDDTSKTSKVYDVNDV-----DTTIEVKDKLGVKTTTLTSTGTGA	1420
Db	1324	DGNLTISK-----SGDSNDVFNLSKDFKVDGTMSTGTTVVND--GVRKGSVDALGTTG	1275
Qy	1421	NKFALSNQATGDALVKAASDIAVHLNLTLSGDIQTAKGASQANSAGYVDADGNKVIYDSTD	1480
Db	1276	LTIANGPAVTASGIDAGSKVISHV-----AAGVASETSDAVNGSQLNAVQVQASQ	1326
Qy	1481	NKYYQAKNDGTVDKTEKVAQKLVQAQTPDGTILAQNNVKSVINKEQ-----	1527
Db	1327	PVTFTEG-NEGAVKRS--LGOSVVISSESTAGTYSGNLKSVVUDEAAGTITHLOLADSPKF	1383
Qy	1528	-----VNDANKKQGI-----NEDNAF-----VKGLEKAAASDNKTKAAVTVG-----DL	1566
Db	1384	GNVVINGGKISGVTAGTETDAVNFSQLKSISTAVDQGWTLTASGANGSKVASGGTVDL	1443
Qy	1567	-NAVAQPTLTFAGDTGTTAKKLGELTLIK-----GGQDTNKLTDNNIGVWAG	1613
Db	1444	KNTDGNLTISKSGDSNDVFNLSKDFKVSSEDLKESITVGNLTQDKDGVKVSNNLLDSNELVI	1503
Qy	1614	TDGF-----TVKLAKDLTNLNSVNAAGTKIDDKGVSVFVDSGQAKANTPVLISANGLDLGK	1669
Db	1504	HSSTSSVKTLTANGESVVRNRTVWNGDGVNIDD--VVVNDLGLSIVGGASLTLSGINAGSH	1561
Qy	1670	VISNVGKTGXTDAANVOQLNEVRNLLGLG-MAGNDNADGNQV--NIADIKK-DPN---	1721
Db	1562	KTNVTAGTETDVAVNFSQLKSVEADVKGWTLTASGANGSKVVSGGTVDLKNTDGNLAI	1621
Qy	1722	SGSSSNRTVI-----KAGTVLGGKGNNDTEKLATGVQVGVVDKDCNANG-----DLSNV	1770
Db	1622	SKSGDSNDVFNLSKDFKDEVTAG-----NTVNTDGVKVSQVSDVSLGAMGLFIANGPSV	1676
Qy	1771	WVKTQDKGSKALLATYNAAQTNYLTNNPA-----EADIRINEQIGRFFHNDGNQEPVVQ	1827
Db	1677	TASGFNAGDK--VISHVAVGMADTDAVNVSQLQAVQSVTVKATRYSTINDGG---TQ	1729
Qy	1828	GRNGIDSSAGKHSVALGFOAKADGAAVAIGHQTOA-GNOSTAIGDNAQATGDQSIATG	1886
Db	1730	GGNYDGDGATGSKAIAAGVGTQASGEGAAVSGGAAAGKGSTAIGRNAIASADGSGVALG	1789
Qy	1887	TG-----NVVAGKHSG-----AIGDPSTVKADNSYSGVNNQFTDAT-----	1923
Db	1790	DGAKDGGRGAEYSTGKYSGVQNTTVGTVSVGDMAKGETRSISNVADAKEMADVNRQLD	1849
Qy	1924	-----QT-----DYFGV--GNNTIVTRES---NSVALGNSAISAGTHAGTQ	1959
Db	1850	AVAQKSNLQTTDDMRHEINNIEDVFKITKGDSSASSVKGMGVNAMAIGTNAAVS-GTESVAL	1908
Qy	1960	AKSDGTAGTITTTAGATGTVKGFAGQAVGAVSVGASGAEIRIIONVAAGEVSATSDAVN	2019
Db	1909	GKNTNVSADNAVAIG-NGSVADRA-----NSVSVSGSGSERQVTNVAAG---TADTDVAVN	1959
Qy	2020	GSOLYKATQCIANATNELDHRIRHONENKANAGTSSAWAMASMPQAVIPGRSMVTGGIATH	2079
Db	1960	VSQI---NQGLITAKOYTDGMVGNLRRTSSGGVAAAIATANLIPQAVYQVQGRGMTSVGVSSY	2016
Qy	2080	NGQGAVALGSLKSDNGQWVFKINGSADTQGHVGAAGVAGPHF	2122
Db	2017	QGQSAVAGVSAVSESGHWVFKEGSGSANTRSHVGVGAGVGYQW	2059
RESULT 4			
ID	Q87DE1	PRELIMINARY;	PRT; 2052 AA.
AC	Q87DE1;		
DT	01-JUN-2003	(T=EMBLrel. 24, Created)	
DT	01-JUN-2003	(T=EMBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(T=EMBLrel. 25, Last annotation update)	
DE	Surface protein.		

```

GN HSF OR PD0744.
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2241331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gaglioti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Camavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
RL EMBL: AE012556; AAC28613.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 4.
DR Pfam; PF05662; HIM; 17.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ
SEQUENCE 2052 AA; 203200 MW; CF194A71FF197655 CRC64;

Query Match 11.5%; Score 1233; DB 16; Length 2052;
Best Local Similarity 23.9%; Pred. No. 2.3e-31;
Matches 583; Conservative 345; Mismatches 779; Indels 730; Gaps 113;

QY 23 AKSHSTGGSCATGCGSVRTLSFARIALAALVIGATLNGSAVAGI--GISEADG--- 76
DB 9 AGTYSGNLKSVDVEAAGAIHLQADSPKFGNVIN---NGGKISGLTAGTDTDAVNLS 65
QY 77 -----GKG-----GANARGKSIAG--DIAQALGSOSIA-IGD-NKIVHNSNNAN 119
DB 66 QLKSIISDVKGWTLTASGANGSKVSGGAVDLAKNTDGNLTISKSGSDNVVFNLSKDFK 125
QY 120 IGAKASG-----NESIAGGDV-----LASGHASIAIGSD--DIYKKETVQOISEL 164
DB 126 VDGMTSGTTVVNNDGVRGVTDAVGTLTGLTIANGPAVTSAGIDAGSKVISHVAAGEVSET 185
QY 165 LPIIRGQKALNDIVQLADT-----NLQYRRTHAQ-----GHASTAVGAMSVAKGHFSN 213
DB 186 STDVAVGSQLNAVQAASQSPVTFSGNEGSKVKTTLGQAVVIGESST---AGTYSGNLKS 242
QY 214 AFGTRATAEGTYSILAV-----GLTATAKAASIIAVG-----SNAQAI 250
DB 243 VVD---EAAGAIHLQADSPKFGNVINNGGKISGLTAGTDTDAVNLSQLKSIISDVTDK 299
QY 251 GFAATAVGGSTQVNLNGIALGFQSGVQLKQNDVNAANVRAYAPDDNQPDNRVK----- 305
DB 300 GWTLTASGAN-----GSKVSGGAVDLKNTDGNLTISK--SGSDNVVFNLSKDFKVD 350
QY 306 -----ATFKNG-----ATDV-----PSIGN-----SNGNDSIRKLIINVAG--SADT 341
DB 351 GMTSGTTVVNNDGVRGVTDAVGTLTGLTIANGPAVTSAGIDAGSKVISHVAAGEVSET 410
QY 342 DAVNVAQLKEAVRLANRQITFKGDDSNRVEKGLGKTLTITGGAGTSALTIDHNTGVVQNG 401
DB 411 DAVNGSQLNAVQAASQSPVTFSGNEGS--VKRTLGAQAVVIGESSTAG-----TVSG 460
QY 402 DGLKVQLAETLTSLNKWTENLTANEKVTVGKTRLTITDKIGFTNDMNGIDESKPYLDKDT 461

```

```

DB 461 GNLSKVDEAAGAIHLQAD-----SPKFG-----NV 487
QY 462 GIHAGGOKITKLTAGVVDDDAATVQOLKVNQTAESALQTFVTKVDK-----NGNDANDS 517
DB 488 VINNGG-KISGLTAGTDTDAVNLSQLKSIISDT-----VDKGWTLTASGANGS 534
QY 518 KIITVGKNNKPDGTQVNTLKLKGNGVDVTTETNG---TVTFLGNQN---NGLTVGNSTL 571
DB 535 KVVSGG-----AVDLKNTDG-NLTISKSGSDNVVFNLSKDFKVDGMTSGTTVV 582
QY 572 NNGLSVONTNSNKKQVQVGDAGITFTDINSKFGAGIENTRITRDGIGFANNTGSLDAN 631
DB 583 NNDGVK-----GSDVALGTTGLTIT-----N 604
QY 632 KPLRTPGINAGKELTN-----VQSAINPATNGGQLDFMNRSLSTANTEKSGSAATIKDL 686
DB 605 GPVATSGIDAGSKVISHVAAGEVSETSDAVNGSQLNAVQAASQ----- 650
QY 687 YNLISOVPLTFAGDTGPNVTKKLGEILKVKGGKTTADDLTAKNIGVAVDSTNSLTVKLAK 746
DB 651 -----PVTFSGNEG-SVKRTLGAQAVVIGESSTAGTYSGNLKSVDVEAAGAIHLQAD 703
QY 747 T-----LSDLDVNTK-----TLTAS-----DKVTVDS 769
DB 704 SPKFGNVINNGGKISGLTAGTDTDAVNLSQLKSIISDTVDKGWTLTASGANGSKV-VSG 762
QY 770 GNNTAKLQNGDLTFKQNTGATPATN-SKTIGVDGLKFTDNNGIALDGTYYITKD--KVG 826
DB 763 GAVDLKNTDGNLTISKSGSDNVVFNLSKDFKVDGM-----TSGTTVVNNDGKVVG 813
QY 827 FAKQDGSIDKSKPYLDKDLKVGE-VEITNGINAGKAIT-----GLSNTLTDATNATT 880
DB 814 -----TDVALGTTGLTIANGPAVTSAGIDAGSKVISHVAAGEVSETSDAVNG-- 861
QY 881 GHVTOLGIVDSTDKTRAASIGDVLNAGFNKNNGDAKDFVSTYDVTDFINGNATTAQVY 940
DB 862 ---SQLNAV-----QAQASQPVTFSG-----NEGSVK---RTLGAQAVVIGESSTAG-TY 904
QY 941 DGKASKVAYDVNVVDGTTIHLTGADGNKNQIGVKTTTLTKTDAKGDKAIFSVNSGDDKAL 1000
DB 905 SGGNLKSVDVEAAG--AIHLQADSPK--FG-----NVLINNG----- 938
QY 1001 INAKDIADNLNTIAGEIRN-TKGTAOT-ALQTFQVKVKGNGDDDDADDTITVKGDAKTN 1058
DB 939 -----GKISGLTAGTDTDAVNLSQLKSI-----SDTVKDWTLTAS 975
QY 1059 QVNTLKLKNGGLDIQTNKDGTVTFGINTOS-----GLKAGNNTLNNGL 1104
DB 976 GANGSKVSGGAVDLK-NTDGNLTISKSGSDNVVFNLSKDFKVDGMTSG-TTVVNDG-- 1032
QY 1105 SIKNTAGNEIQVQADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFAGTNGSLDKSKPHLS 1164
DB 1033 -----VKVGS-----VALGTTGLTI-----TNG-----PAVT 1055
QY 1165 KGINAGGKITTINQSIEIAQNSNDVATGKIYDLKTELENKISSAKTAKQNSLHFSVA 1224
DB 1056 ASGIDAGSKVISHVAAGEVSETSDAVNGSQL-----NAVOAQA----- 1094
QY 1225 DEQNNFTVSNPYSSYDTSKTSVDITFAGENG-IITTKVNGVVRVGIDQTKGLTTPKLT 1283
DB 1095 -----SQPVTFSGNEGSVKRTLGAQAVVIGESSTAG-----TY 1127
QY 1284 GNNNGKIVIDSQNGQNTITGLSNTLANVTNDKGSVRTTEGKI-----IKEDK--T 1334
DB 1128 SGNLKSVDVEAAGAIHLQADSPKFGNVLINNG-----GKISGLTAGTDTDAVNLS 1180
QY 1335 RAASIVDLVLSAGFNLQNGE-----AVDFVSTYDVTNFADGNATTAQVTDYDTSKT 1385
DB 1181 QLKSIISDVKGWTLTASGANGSKVSGGAVDLKNT-----DGNLTISK-----SGDS 1258
QY 1386 SKVVYDVND-----DITIEVKDKKLGKTKTLTSTGTGANKFALSNOATGDALVKA 1437
DB 1229 NDVVFNLSKDFKVDGMTSGTTVVND---GVKVGSDVALGTTGLTIANGPAVTSAGIDAG 1285

```


Db 1440 LOGANGTVGLTGADGAQAVITVYKGRPTLDNAEPTPIAAGNEEVATLNDGLKFGNAGD 1499
QY 903 VLNAGPNLKNNGDAKDFVSTYDTVD-----FINGNATT-----AKVTYDGKAS 945
Db 1500 VHNAKLNTQVDYKGAATANTVWDFKQGNIMTRVEGNTITVALAKALSGLTSATFGDPAS 1559
QY 946 --KVAYDVNVVG-----TTIHLT--GADGNKNOI-----GVKTTTLTKTDAKDKALN 989
Db 1560 NPKDSVTINOKGLTTTQGNONTVSLTDDGLDNGKQIKNVASGLTTNGTATSLDDAVOT 1619
QY 990 FSVNSGDDKALINAKDIADNLNLTAGEIRNTKGTADTALQTFQVKVKVKGNGDDNDADTI 1049
Db 1620 NGVNVGDLKTAI--NITNGTPLGG-----FGLKDKAGNTEK-----QNLGETA 1662
QY 1050 TVGKDAKNTQVNTLKLKGN--GLDTQNKDGTVTFGINTQSGLKAGNNTLNNGLSIKN 1108
Db 1663 QITGD---SNVNTKVVDGQGGKALEVSLANQLTLGKGPANVP---NATGAGKITLKD 1716
QY 1109 TAGNEQIQV--GADG-----VKFAKVNNGVVGAGIDGT-----TR 1141
Db 1717 DKGTRVVVDGEGALSILTGQATQGAAPTAKIKVAGNPDLENTSDDPANNQNKTR 1776
QY 1142 ITRDEIFAGTNGSLDKSPHLSKDGINAGGKKITNIQSGEIAQNSNDVATGKIYDLKT 1201
Db 1777 ITYD---IAGPNGTIVTEQLATLNDGLKFGA-----NTGDV---HDAKLNTRV-DYKG 1822
QY 1202 ELENKISSTAKTAQNSLHSEFVSADQGNFTVS-----NPYSYDTSKTSDD 1247
Db 1823 KAENTNWFADAGQNTMTQIS-----GNTITVALAKALAGLDSATFGNP---ADGSKDGA 1874
QY 1248 VITFAGENGITTKVNVGVVRGIDOTKGLTTPKLT--VGNNGKGVVIDSQNGTITGLS 1306
Db 1875 VI-----NND-----GLTTTEGDKTVKLTTEKGLDNGGQNLINVDLSGLKKTGDG-- 1916
QY 1307 NTLAVNTNDKGSVRTTEQKIIKDEKTKRAASIVDVLSA---GFNLQNGEAVDFVSTYD 1363
Db 1917 -SVVALKDAEGSVLT--NGVNVGDLKN---AIKDVTSATNGFGGLKDKAGA----- 1961
QY 1364 TVNFDAGNATTAKVYVDTSKTSKVYVNVVDVTIEVKDKLGVKVTTLTSTGTGANKF 1423
Db 1962 --EFKQDLGTTTAQITGDKNINT--KVIVDPVNSDKALEI----- 1996
QY 1424 ALSNQATGDALVKASDIVAHLNLTSGDIQTAK--GASQANSSAGYVDADGNKVIVYSTDNK 1482
Db 1997 -----SLANDITLGRKNGADGVDSGLVNGKOGASVVLNG----- 2030
QY 1483 YYQAKNDGTVDKTEKAVDKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANKKQGINEQNA 1542
Db 2031 -----KDGSGIGLTGPRGD-----GSDGKSATISVK-----DGRAGVDGKD- 2066
QY 1543 FVKGLEKAASDNKTNA--AVTVGDLNVAQTPLTTFAGDTG--TTAKLGETLTKGG----- 1596
Db 2067 ---GDTKTRIVYETKDATKPVVEEATLNDGMKFVNGDKREVTRKLNETLIDKGLDAA 2123
QY 1597 -QTDNKLTDNNIG-----VVAGTDG--FTVKLAKD----- 1624
Db 2124 TVADNAKVSSSLGVLKVTNAEGTGLEIWMKERPTFSLVWNGKDGEDAAVKFARKDGKQMS 2183
QY 1625 ---LNLMSVNAAGGTIKIDK---GVSFVDSGQAKANTPVLSSANGLDGGKVISNVGKG 1677
Db 2184 IAAVTD--NDGNATGLTIKDKGNPGVTF--NNDR-----ITNVTAG 2222
QY 1678 TKDTPAANVOQLNEVRNLLGLGNAGNDNADGNQVNI--DIKKDPNSGSS-----N 1727
Db 2223 VDDKDAVNVSQLD-----GLAKATTKVEAGKNMTVPTVND---GSTTYTVATEDNVN 2274
QY 1728 RTVIRAG--TVLGGKGNNDTEK-----LATGGVQGVGDK--DGNANGDL 1767
Db 2275 FTTVTGTNTW---NNDGVKVGDNVNLNEGKAGDVTVTTAGINAGNKKYTGVDAGDI 2330
QY 1768 S-----NVWVKTKQKSGSKALLATYNAAGQTNVLTNNPAAIDRINEQG--IRFFHVN 1818

Db 2331 SPNSTDAVNSQLNAVKETABAGWHLTANGADSSNV---KPRNTVDLNNTDGNIVISKTN 2387
QY 1819 DGNQPEVVOGRNGIDISSAGSKHSVAIGFOAK--AGEAAVAIGRQTOA--GNOSIALGDNA 1875
Db 2388 TADKENVT---FGLADNINVKDSSVVVPGKANGKFGEGAWVNAEDGANGKDGISI---V 2441
QY 1876 QATGQOSIAI---GTGNV-----VAGKHSGA--IGDPSTVKADNSYSVGNNNQFTDAT 1923
Db 2442 GKGDKDAVALSKDGVGTIIGLTPAGADGNKNAIIGVNDVKGLDNGDKGNSKTRIV 2501
QY 1924 QTDVFGVGNNTITVTSNSVALGNSAISAGTHAGTQAK--KSDGTAGTITTTAGATGTVKGF 1982
Db 2502 YTKPNGEEOV--ATWMDGLVFCDKGTBHKALGTTTVKVGDDKNIEVEAGDTIRV--L 2559
QY 1983 AQQTAVGAVSVGAS-----GAERRIQN-----VAAGEVSATSTDVANGSOLYKATQGI-- 2030
Db 2560 KDNIDVKGINVTENLTVKEGAKINGNVNDVGDVADGEVNAVTSKQAVNGSQLHKVQQOVNN 2619
QY 2031 -ANATNELDHRTHQNEKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAVGL 2089
Db 2620 QATAINKLDHINKVDKDLRAGIAGATAVAFQRPNEAGKSVLSLVGVSYSRSESAIAVGY 2679
QY 2090 SKLSDNGQVFPKINGSADTQGHV--GAAVG 2117
Db 2680 ARNSDNNKISIKLGGGMSNRGDNVFGSIG 2709
RESULT 8
Q8XDG4 PRELIMINARY; PRT; 1588 AA.
AC Q8XDG4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative adhesin (Hypothetical protein).
GN Z5029 OR ECS4480 OR H161.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 0509952;
RA Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,
RA Shinagawa H.;
RT "O157 specific gene similar to H. influenzae adhesin gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE005587; AAG58749.1; -;
DR EMBL; AP002566; BAB37903.1; -;
DR EMBL; AB036416; BAB7814.1; -;


```

QY 1877 ATGDOSIAIGTGMVAGHSGAIGDPSTVKADNSYSGVNNQFTDATQTDVFGVGNITV 1936
Db 1388 -----SVALGTGSV-----ATENTISVGSSTNQ 1411
QY 1937 TESNSVALGNSAISAGTHAGTQAK-KSDGTAGTITTAGATGTVKGFGAGTAVGAVSVGA 1995
Db 1412 RRTITVAAGKNA-----TDAVNVQALKSSEAGGVRYDTKADGSID--YSNITLG-----G3 1460
QY 1996 SGAERRIQNVAAGEVSATSDAVNGSOLYKATQIGIANATN-----ELDHRIHQENKANAG 2051
Db 1461 NGGTRISNVASG--VNNNDVVNYAQLKQSVQETKQYTDQRVEMDNKLSKTESKLSG 1517
QY 2052 ISSAMAMAPQAYITPGRSMVTGGIATHNGOGAVAVGLSKLSDNGOWFEKINGSADTQGH 2111
Db 1518 IASAMAMTCLPQAYITPGASWASIGGTYNGESVALGCVMSVANGRWYKLGSTNSQGE 1577
QY 2112 VGAAGVAGPHF 2122
Db 1578 YSAALGAGIQW 1588

RESULT 9
Q8GM77 PRELIMINARY; PRT; 1004 AA.
AC Q8GM77;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Adhesin.
GN HIA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=montypable strain 1862A;
RX MEDLINE=22300350; PubMed=12410830;
RA Laarmann S., Cutter D., Juehne T., Barenkamp S.J., St Geme J.W.;
RT "The Haemophilus influenzae Hia autotransporter harbours two adhesive
RT pockets that reside in the passenger domain and recognize the same
RT host cell receptor.";
RL Mol. Microbiol. 46:731-743(2002).
DR EMBL; AY078086; AAL79952.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR Pfam; PF05658; Hep_Hag; 2.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 1004 AA; 104286 MW; 7806E7E2D4E3D017 CRC64;

Query Match
Best Local Similarity 29.2%; Pred. No. 1.5e-22;
Matches 335; Conservative 142; Mismatches 337; Indels 332; Gaps 47;

QY 1124 PAKYVNGVVGAGIDGTTTRITRDEIGFAGTNG-----SLDKSPHLKSDGINAGKK 1173
Db 44 FATVQASV--AATSGTT-----GTNGLHTYGDASFVFNNSATDLNRH-VEDAYK 90
QY 1174 KITNI-QSGEIAQN-----SNDAVTGKIKYDLKTELENKISSTAKTQNSLHEFSVADEQ 1228
Db 91 GLILNLEXGADKSNFLVANDTTATVGNLRLKGLWLSKKNGTRNEKSQQVKHAEVLFEK 150
QY 1229 NNFTVSNPYSSYDTSKTD---VITPAGENGITTKVKNKGVVRVIGIDQTKGLTTPKLTG 1285
Db 151 DGVTV-----TSKSENGKHTVTFTLANDLVK-----NATVSDKLSLG- 188
QY 1286 NNGKGVIVDSQ-----NGQNTITGLNLTANVTNDKSGVTTTQGGKLIIDEDKT 1334
Db 189 ANGKKVDITSDANGLKFAKPGTNGQNGVHLNGIASTLDDPRVGGKTAHLTKETISDTERN 248

```

```

QY 1335 RAASIVDLVSAGFNLO-----GNGEAVDFVSTYDTNFAENATTAKVTYDDTSKTSKV 1388
Db 249 RAASVGDVLNAGNWRIGAKTIGTVDNVDFVSTYDYTFEASGANANVSVTTDDNKKT-- 306
QY 1389 VYDVNVDDTTIEVK-----DKKLG-----VKTTTLITSTG 1417
Db 307 ---VRVDVTGLPVQVVTEDSKTVKVGNEYEYEAQDSADMCKKVENGLAKTKVKLVSA 363
QY 1418 TGANKFALSNOATGALVKASDIVA--HLNTLSGDIOTAKGASOANSAGYVDADGNKVI 1475
Db 364 NGTNPVKISNADG--TEDTDAVSFKQLKALQ-DKQVTLASANAYANGG-SDADGGKVT 418
QY 1476 -----YDSTDNKY--YQAKND-----GTVDKTKVAKDKLVAQATPDGTLA 1515
Db 419 KTLNNGLNPKFKSTDCGLLNIAKENDTVTFPKKGSV-QVGGDGGKASISKGANTIEGLVE 477
QY 1516 QMNVKSVINKEQVNDANKKQGINEDNAFVKLEKAASDNKTKNAAVTVVGDLANAVAQTPLT 1575
Db 478 ASELVDSLK-----LGWKVGVDKDS-----GELDGTPEKETLV 511
QY 1576 PAGDTGTTAKKLGETLTIKGGQDTNKLTDNNGVAGTDTGTVKLAKDLTNLSV----- 1631
Db 512 KSGDKVTL--KAGDNLKVK--QEGTN-----FTYALRDELTVGKSVFEKD 552
QY 1632 -----NAGTKLIDDKGVSPVDSG-----QAKANTPVLISANGLDLGGKVISNVGKTQDT 1681
Db 553 TANGANGASTKITKDLTITPANGANGAATDADKIKVASDGI.SAGNKAVKVVSLKKF 612
QY 1682 DAANVOQLNEVRNLLGLGNAGNDAGNOVNIADIKKDPNSGSSSNRTVIKAGTVLGGKG 1741
Db 613 GDANF-----DPLTSSADNLT-----KQ 630
QY 1742 NNDTEKLATGGVQGVGDK-----DGNAN--GDLSNV-WVKTKQDGSKALLATYNAAGQ 1792
Db 631 NDDAYKGLTNLDEKGTDKQTPTVAONTAATVADGLRLGWV--ISADKTTGELNKEYNA--- 686
QY 1793 TNYLTNNPAEADIRINEQGIREFHVDNGQEPVVOGRNGIDGS---ASGKHSVAIGFOAK 1849
Db 687 -----QVRNANEVFKSG-NGIHVSGKTVNGRREIT--FELA 720
QY 1850 ADGEAAVAIGROTAQGNQSIAGIDNAQATGDSIAIGTGNVNVAGHSGAIGDPSTV--KA 1907
Db 721 KDENA-----IAFGYGSKALRDNTVAITGTVNVVNAEKSGAFGDPNYIEDKA 766
QY 1908 DNSYSVGNNOFTDATQTDVFGVGN-----ITV----- 1936
Db 767 GGSYAFGNDNRIT-SKNTFVLGNGVNAKYKANGVDVTETVTVKDKDGKETTVPVKALGA 825
QY 1937 TESNSVALGNSAISAGTHAGTQAKKSDGTAGTITTAGATGTVKGFGAGTAVGAVSVGAS 1996
Db 826 TVNSVYLGNK---STATKDKGNLKSDDGTAGNTTITAGTITGTVNGFAGATAHGAVSUGAS 882
QY 1997 GAERRIQNVAAGEVSATSDAVNGSOLYKATQIGIANATNELDHRHQENKANAGISSAM 2056
Db 883 GEERRIQNVAAGEISATSDAINGSOLYAVAKGVIN---LAGQVKNVKRADADAGTASAL 938
QY 2057 AMASHPQAYIPCRSMVTGGIATHNGOGAVAVGLSKLSDNGQVFKINGSADTQGHVGA 2116
Db 939 AASQLPQASMPGKSMVSIAGSSYQGNGLAIGVSRISDNGKVIIRLSGTTNSQGGKTGVA 998
QY 2117 GAGPHF 2122
Db 999 GVGYQW 1004

RESULT 10
Q8GM78 PRELIMINARY; PRT; 1002 AA.
AC Q8GM78;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Adhesin.

```


QY 1480 DNKYVQAKNDGTVKTEVAKDKLVAQAQTPDGTLAQMNKSVINKEQVNDANKKOGINE 1539
 Db 911 ----- 910
 QY 1540 DNFAVKGLEKAAADNKTNAATVTVGDLNAVAQTPLTFAGDTGTAKKLGETLTIKGGQTD 1599
 Db 911 -----VTVG-----SGVTLSGMLVI----- 926
 QY 1600 TNKLTNNLGVVAGTDTFTVKLAKDLTNLSNVNAGTKIDDKGVSVFVDSGQAKANTPV 1659
 Db 927 -----TDG-----PSV 932
 QY 1660 SANGLDLGGKVISNVGKGTDDAANVQOLNEVRNLLGLGNAGNDNADGNQVNIADIKD 1719
 Db 933 TSSGINAGSQKTINVAAGTADTDVNLSQLN----- 963
 QY 1720 PNSGSSNRVTIKAGTVLGGKGNNDTEKLAT--GGVQGVVDKDGNGANGDLSNVVWTKQD 1777
 Db 964 -----TAMASSGAKSVMHYSTYDGGTQ-----CGNYNGD----- 992
 QY 1778 GSKKALLATYNAGQNTYLTNNPAEALDRINEGIRFFHVDNQGPEPVQGRNGIDSSAS 1837
 Db 993 -----GAT 995
 QY 1838 GKHSVAIGFOAKADGEAAVAIGRQTOA--GNQSTAIAGDUAQATGDSIAIGTG----- 1888
 Db 996 GTGSTAVGVGTILASAGATAIGSAAASGKSTAIGRNAVASADGSVALGDGAKDGARGA 1055
 QY 1889 NVVAGKHSG-----AIGDPSTVKADNSYSVGNNOFTDAT---QTDVFG-----V 1930
 Db 1056 ESYTGKYSGVQNTVGTVSGVDASKGETRTVSNVADAKEATDAVNLRLQDLQVAQDANRV 1115
 QY 1931 GNNI-----TVTESNVALGSNSA--ISAGTHA-----GTQAKKSDGTA--GTTTTAGA 1975
 Db 1116 DNKIESLSEGTQFVKVNSV---NNSATPIAAGVDATAIGVGATAGSADSIAMGNKASASA 1172
 QY 1976 TGTVKGFAQGTAVG-----AVSVGASGAERIRIONVAAGVSVATSSTDVANGSOLYKA 2026
 Db 1173 DNVA-----AIGHSVADRANTVSVGSAGSERQVTINVAAG---TADTDVNVSQL----- 1219
 QY 2027 TQGIANATNELDRIHQNEKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVA 2086
 Db 1220 NQGLITAKQYTDGVVGLRRDRTDGGVAAAIATANLPQAYIPGRMTSVGVSVYRGOSA 1279
 QY 2087 VGLSKLSDGQWVFKINGSADTQGHVGAAGVAGFHF 2122
 Db 1280 VGVSVSVESGRWVFKFSGSANTRSQVIGAGVGYOW 1315

RESULT 13

Q98HJ2
 ID Q98HJ2 PRELIMINARY; PRT; 1953 AA.
 AC Q98HJ2;
 DT 01-OCT-2001 (TremBLrel. 18, Created)
 DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Hypothetical protein ml12848.
 GN ML12848.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RC MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti "1";
 RL DNA Res..7:331-338(2000).
 DR EMBL; AF003000; BAB49874.1; --
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR000911; Ribosomal_L11.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep_Hag; 30.
 DR Pfam; PF05662; HIM; 8.
 DR Pfam; PF03895; Yada; 1.
 DR PROSITE; PS00359; RIBOSOMAL_L11; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;
 Query Match 7.8%; Score 831; DB 16; Length 1953;
 Best Local Similarity 22.5%; Pred. No. 1.3e-18;
 Matches 540; Conservative 251; Mismatches 828; Indels 780; Gaps 107;
 QY 17 MAVAEYA-----KSHSTGGSCATQGV-----GSVRTLSPARIALAVLVIGAT 60
 Db 39 MAVLRMAHRLVGMRRKSGSLGMLGIRVLAALGATLGLGIATPALAQYAA-----GGGT 93
 QY 61 LNSAYAGIGISEADGG-----KGGANARGDKSIAIGDIAO-----ALGQSQSTAI 107
 Db 94 ANSAGSVAVGPSATTNGLRGIAGVSGATQAGIDISIAQGTSAAGSQAQVVAIGFQSIASQL 153
 QY 108 NKIVHNSNNANIGAKASGNESIAIGDVLASGHASIAIGSDDLYLKKTETVQ--QISELLP 166
 Db 154 NSIYIGSRTAAGTGATA--QSAIGIGTDTVASQVDALAVGRSSVASAQYSVALGJLSAKAT 211
 QY 167 IIRGOKALNDIYQLADTN--LQKYRTHAQCHASTAVGAMSAKGFHFSNAGFTRATAEGTY 225
 Db 212 GTGGAWALQGGTIANINSVALGVQASATGGCANALGTFVSAGGNGSTAVGTSSSAAGSL 271
 QY 226 SLAVGLTATAKAASSIAGVSNQAQIGFAATAVGGSTOVNLRGIALGFQSGVQLKQNDVN 285
 Db 272 SPAGGWSAVASGDSVALGKSANAAGLNSALGSGSTNASADPALAL----- 317
 QY 286 AANVRAYAPDDNQPIDNRYKATFKNGATDVFISIGNSGNDSIRRKTIINVAGSADTDVAVN 345
 Db 318 -----GNOAVSSGIGSVAAGSG--SQATG 339
 QY 346 VAQLKEAVRLANRQITFKGDDSNRVERKGLGKTLTITGGAQTSALTDRHNVGVVQNGDGLK 405
 Db 340 VS-----ATALGNNA---TAANATALGLGAT---AGGVDCGA-----IGKQAN----- 378
 QY 406 VOLAETLTSLKMTVTENLTANEKVTVGKRLTTD-----KIGFTNDMNG---IDESKPYLD 458
 Db 379 ---ASAQDAIAMGTSAKASSAQAIAGVINAATGKAVSIGSGNTAYGDGAVSIGDPSYA 435
 QY 459 KDTGIHAGGQKITKLTAGVWDDDAATYQQLKKVNTAESALQTFVTKVKVKNKNGNDANDSK 518
 Db 436 SGTGAFTGANNIANSDDGTATATA-----NOAAGA----- 466
 QY 519 IITVGKNNKP-----DGTQVNTLKLKGE-----NGVDVTETNGVTFTGLNQNNGLT 565
 Db 467 -VAIGNNKAIGQGSVALGNGSTAGAGLAGNVALNGA--TAAASGSDVALGSGSVTAVA 524
 QY 566 VG--NSTLNNGLSVKNTNSNKQIQGVADGI--TFTDISNKPAGKAGINTTRITRDGIGF 621
 Db 525 VGTPTNAVINGTITAFQGTPTPTSTVSIGAPGAERTLTNVA-----AG-----RIS----- 568
 QY 622 ANNTGSLDANKERLPTPTGINAGKELTNVQSAINPATNGSQLDFMRLSTANTEKSGSNA 681
 Db 569 GSSTDAVNGSQLFATNQAVDAIGTTLNNIN-----GGGIKYPH-----ANSTLADSSA 617
 QY 682 TIKDLNLSQVPLTPAGDTGNVTKLGEILKVKGGKTTADDLTQKNIGV--VADSTONS 739
 Db 618 TGTDSVAI-----GPTSTATASSAI--AGSNANASCANASAI--GTSVSAALDAT 665

```
QY 740 LTVKLAKTLDLDAVNTKTLTASDKVTVDSNNNTAKLQNGDLTFPSKO--NTGATPATNSKT 798
Db 666 AMGFSLKASQSFSTAVGANANATLSTALQNALASGVQATALGKQANAGA-----SEA 720
QY 799 IGVNGLKFTDNNGLDGTYYITKDKYGFARQDGLSKRPLDKKLKVGVEITNGI 858
Db 721 LALGANSTAGNAGDVALGSGSVTAVAG-----TPNAVINGTYYAFQGI 764
QY 859 NAGKAITGLSNTLTATNATTHVHTOLGIVDSIDTKTRAASI---GDVLNAGFNKNGD 915
Db 765 NPASTVSIAGPAGERTLTNLAAGRISG-----SSTDVANGSOLFATNOAVDAIGTVNN-- 818
QY 916 AKDFVSYDVTDFINGNATTAKVTYDGKASKVAVDVNVVDGTTHLTGADGNKQIG-VKT 974
Db 819 ----ISTGGGIKYFHANSTLA---DSSA-----TGTD--SVAIGPTST 852
QY 975 TLTUTKDAKDAKAINFSVNGDDKALINAKDIADNLN--TLAGEIRNTKGTADTALQTFQV 1033
Db 853 ATASSAIAAGSNA-----NASCANASAICTSSVASALDATAMGFLSKASQGFSTA----- 902
QY 1034 KVKENGDDDDADTTITVGDAKTNQVNTLKLKKNGLDITQNKDGTVTTCINTQSLKA 1093
Db 903 -----VGANANATLSTAI-GQNAL-----ASGVQATALGKQANASA 939
QY 1094 GNNNTLNNGLSIKNTAGNE-QIQVGADGVKPAKVNNGVVGAGIDGTTRITRDEIGFAGT 1152
Db 940 SDALALGAN-----STAGNAGDVALGSGSVTAVAG--GTPNAVINGTIT-----YAFQGI 986
QY 1153 NGSLDKSKPLSKDGINAGG--KKITNIQSGEIAQNSNDVATGGIKYIDLKTELE----- 1204
Db 987 NPA-----STVSIAGPAERTLTNLAAGRISGSTDVANGSOLFATNOAVDAIGTV 1038
QY 1205 NKIS-----STAKTQNSLHEESVADQGNFT-- 1232
Db 1039 NNISTGGGIKYFHANSTLADSATGTDSVAIGPTSTA-TASSAIAAGSNAANASAI 1097
QY 1233 -VSNPYSSYDT-----SKTS-DVITPAGENG-----ITTKVNGVVRVVGID-----QT 1273
Db 1098 GTSVASALDATAMGFLSKASQGFSTAVGANANATLSTALQNALASGVQATALGKQA 1157
QY 1274 KGLTTPKLTGVNNGKIVIDSQNGNTIT-----GLSNTLANVT-----N 1314
Db 1158 NAGASDALALGANSTAGNAGDVALGSGSVTAVAGTTPNAVINGTYYAFQGINPASTVISG 1217
QY 1315 DKGSVRTEQ---GKLIKEDKTPAASIVDLSAGFNQNGEAVDFSVTYDVTNFDGN 1371
Db 1218 APGAERTLTNLAAGR-----SGSSTDVAVN-GSOLFATNOAVDAIGT--TVNNLGGG 1266
QY 1372 ATT-AKVYDDTSKTSKVVDVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNOAT 1430
Db 1267 VTNLGNVNNIAGDTSYATDANGIGIRY-ARTNEAGLAQTDSPAQGLGSTAVGYQASAT 1325
QY 1431 GD---ALVKASDIVAHLNLTSGDIQTAKG--ASQANSSA-GYVDA-DGNKVIYDSTNKY 1483
Db 1326 GVSGLALGRGT-----LASIDGVSALGSGSVADRAVPATGQIAAGPANFIQYNTDKTL 1380
QY 1484 YQAKNDGTQVTKTEKAVDKLVAQAQTPDGTGLAQNMVKSVINKEOVNDANKQGINEDNAF 1543
Db 1381 LGAVSVGTATSYRQI-----TNVADGTQAO-----DAVTVRO----- 1412
QY 1544 VKGLEKASDNKTK--NAAVTGVGLNNAVAQTPLTTFAGDTGTAKKLGETLTIKGGQTDTN 1601
Db 1413 LOGVIASVSATSTKYFHANSTAGDSLAV-----GAESVAVGPTTVWNG----- 1455
QY 1502 KLTDNNIGVAVG-----TDGPF---TVKLAKDL-----T 1626
Db 1456 ---DNGTIGINGALVDAQPGTGAIGNAHVLMADGLALGTNSLASIGSVALGAGAQT 1512
QY 1627 NLNSVWAGGTIDDKGVSFVDSGQAKANT-PVLSANGDLIGG----KVISNVGKGTQOT 1681
Db 1513 FINSVALGAQSVTVVG---AQAGYATALTAPQTSVGEVSIIGGAGAEKRLTNVAAGSANT 1569
```

```
QY 1682 DAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPNSGSSSNETVIKAGTVLQCKG 1741
Db 1570 DAVNVSQLK-----GVSNVANL-----FGG-- 1590
QY 1742 NNDEKLTATGGVQGVYDKQNGANGDLSNWWKTKQDKGSKALLATYNAAGQTNVLT---- 1797
Db 1591 -----STTVNPDGSI TG-----PTYTIQG--NNYSTVYDG 1618
QY 1798 ---NNPAEAI DRINEQGIREFPHVNDGNQBPVVQGRNGIDSSASGKHSVAIGFOAKADGE 1853
Db 1619 FTAVNNALTNIS--NGGGIKYFHANS-----TLADSTAGTDSVAL-----GP 1659
QY 1854 AAVAIGRQTOAGNQSGIAIGNDAQATQDSIAITGTGNVYVAGKHGAIGDPSTVKADNSYSV 1913
Db 1660 ASVA-----SGTNSLAAGNGSTATQGAVALGQ-----AKANN----- 1693
QY 1914 GNNNQFTDATQTDVFGVGNNTITVTSNSVALGSNSAISAGTHAGTQAKKSDGTGATTITA 1973
Db 1694 -----ANDVALGSGSV-----TQ-----TAVGTST 1714
QY 1974 GATGTVKGFAGQTAGVAVSVGASGAERRIQNVAAAGEVSATSTDAVNGSOLY----- 2024
Db 1715 VINGKYTAPAGTPTGTIVSVGDAGAERTITNVAAGRVNAGSTDAINGSQLYATNATVEDL 1774
QY 2025 ---KATQ-----GIA-----NAT---NE 2036
Db 1775 KSLGLSLTONAVTVDTNPDGSKKNSITLQGGDVNAPVVISNVGPGVAGTDAVNVNOMNR 1834
QY 2037 LDHRIHQN-----ENKANAGISSAMAMASMPQ 2063
Db 1835 VDIATNITSNYTDKVAATTLQOANNVTDOKLSQNSLDGIRDEARQAAIGLAASLRY 1894
QY 2064 AYTPGRSMVTGGIATTHNGQGAVALGSLKSDNGQWVFKINGSADTQGHVGAAGVAGGPHF 2122
Db 1895 DDRPKLSVAAAGGGFWRDSALAFAGYGTSEDGEIRGNVSGTA-AGGHVG--VGAGISF 1950

RESULT 14
Q8PCQ5 PRELIMINARY; PRT; 2351 AA.
AC Q8PCQ5, 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DR 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XCC0658
OS Xanthomonas campestris (pv. campestris)
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C. de Souza R.F.,
RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DB EMBL; AB012164; AM39974.1; -.
```

```

DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 59.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2351 AA; 220261 MW; C5482A38C940DA11 CRC64;

Query Match      7.3%; Score 784.5; DB 16; Length 2351;
Best Local Similarity 20.7%; Pred. No. 4.8e-17;
Matches 556; Conservative 297; Mismatches 936; Indels 891; Gaps 104;

QY 1 MNHIYKVIENKATGTFMAVAEYAKSHSTGGSCATQGVSVRTLSPARTAAALAV-----54
Db 1 MNRIYKWNKSLGWAVASSELASGDSPGSVASA-ALIDRRQGLSLAAATLALGAGIA 59
QY 55 -----LVIGATLNGSAYAGIGISEADGKGGANA-----83
Db 60 IPLSASAQSVVGRGASAPAAHATAIGAGSNASA-TGAVATGADSSASGVNSAIGRQTN 118
QY 84 -----RGKSTAIADIAQALGQSIAIGDNKIVH-----112
Db 119 AIGENAVAIYNSFVRQAGENGVALGANAGVTGANSVALGAGSRTHEDDVVSVSGNGRG 178
QY 113 -----NENNAN-----IGAKASGNESIAI 132
Db 179 GPATRRITNVSGVNVANDVNVQAQVSEVABDTATFFKAQPGDSDSIGAYADGAGAVAA 238
QY 133 GGDVLASGHASIAIGSDDIYLKKTETQOISELLPIIRGOKALNDIYQL-----180
Db 239 GDAANAFSGALAVGSGANALADNATAMGANALAVGNSAAFGYNAQAQPGSVAVGGAA 298
QY 181 -----ADTNLQKY-----RTHAQGHASTAVGAMSYAKGHFSNAPGT 217
Db 299 VDEGNGPLITNGVPVYDTGATSGVGGTAVGASASADGFAASVYGVGAAGAQAQSAFGV 358
QY 218 RATAECTYSLAVGLTATAKAASIAVGS-----NAQAIGFAATAVGSGTQVNL 265
Db 359 VSNATGDYTAGVTQTSAGTSTSTAVGGPVLDLIPGLGFFVQTQASGEAASALGAGAIASG 418
QY 266 NRGIALGFGSQVLQKNDVNVANVRAYAPDNPIDN-----RYKATPKN--310
Db 419 TYTTAVGTLSEA--SGTEATVGYFAYAPCEGATAVPESWASGELSTALGYSTARGAN 476
QY 311 -----GATDYFSTGN--SNGNDSIRRKIIINGGASADTDVNVQAQLKEAVRLAN---RQIT 361
Db 477 SVALGANSVATRADTVSVGAAGAEROITSVAACTEGTDAVNLNQLTAVSDVASATARSFV 536
QY 362 FKGDSDNNVEKGLGKTLATITGGAQTSALTDDHNIQVGVQNGDGLKVLQALETLSLKWTTTE 421
Db 537 ATGDGA--AIAEGVD--SVAAGSNAAFNDYST-ALGSSSVASAGATAVGSGANATTD 590
QY 422 NLTA-----NEKTV-----GKTRLITD-----KIGFTNDMNGIDESK 454
Db 591 NATAVGFNSTVAENTTALGNSASSGDSGSTAVGGATRATASCATALGYESIANGADSTA 650
QY 455 PYLDK-----DTGIHAGGQKITKLTAGVDDDAATYQOLKKVNOTASALQTFVTKKVDKN 510
Db 651 LGVGSVAFGGTSTAVGGSVAFGT-----DAAFGANAAAGGTASTA-----IGAN 696
QY 511 GNDANDSKLITGVKNKPKDQTVNTLKLKENGVDVTEINGVTVFLNQN-----NGLTV 566
Db 697 SN-AGERTVALGGASNAGDESIAL-----GVSSIASALGTTAVGSNNANASIANATAV 749
QY 567 G-NSTLNNND-----GLSVKNTNSNKQIQVGADGITFTDISNKP 605
Db 750 GFNSAGDDVATALGDSNASGYFSTAVGTSIANGR-----GATAIGYESIINGAAS 802
QY 606 AGI-----ENTTITRDIGIFANNVGSGLDANKERLPTPTGINAGGKELTNVQSAINPA 657
Db 803 TALGFASVAMVEGGTAIGTESVAYGDDSTALGAN-----836
658 TNGGQLDFWNLSTANTEKSGSAATIKLDYLNLSQVPLTEAGDTGPNVTKLGLBILKVGK 717
837 -----AYAADTGSIAV-----GTAYANAYGPRAIS-----LGG 863
718 KTTADDLTKNNTIGVADSDNSLTVKLAKTLSLDADVNTKTLTASDK-VTVDSGNNATKL 776
864 QSNA-----AGDESI-----ALGWEAQAEQDGIAGAGSQAD--896
777 QNGDLTFSQNTGATPATNSKITIGVDGLKFTDNGIALDGTITYITKDKVGFQAKQDGLDX 836
897 -----AYSTAIGYATASGATAV-----GNNSRAVDG--YATA-----929
837 SKPYLDKDKLKVGEVEITTINGIN-AGGKAITGLSNLTLDATNATTHVTLQGLIVDSTDKT 895
930 -----LGSDSKASGNFSTTVGGASVASGRGATAGAESVARMDBDITAIGTE-SVADGSDST 984
896 RAASIGDVLNAGFNKNNGDAKDFVSTYDTPFINGNATTAKVYDGRKASKVAYDVNVDG 955
985 A-----LGNAR-----ASYDSVALGANANSSNYV--SVALGTYAVATGG 1023
956 TTIHLTG---ADGNKH-QIGVKTTTLTKTDAGDKDAINFVNSG-----DDKALINAKDI 1006
1024 SATSIGQGSYAFGNESVALG-----WQSNASGTRSV--SLGSGAYTPADDDGVALGAGSI 1075
1007 ADNLNTLA-GEIRNTKGTADTALQTFQVKVKKENGDDDDADTI-TVGKDAKTNOVNTLK 1064
1076 ARDNTVSVGSVSEROITNVAAGT-----EGTDAVNLQNLNAVAGTAET---TAR 1123
1065 LKQKNGLDIQTKNDGTVTFGINTQSGLKAGNNTTLNNN-----GLSIKNTAGNEIQVGA 1119
1124 LVAGTG-----DGTAF-AQGEDATAAGSNATADADYSSAFGASSQATA-----IGA 1168
1120 DGVKFAKVNNGVAGIDGTITRDEIGFAGTNGSLDKSKPHLSKXDGINAGGKKTINIQ 1179
1169 -----VAIGSGASATAQ-----YANASGYNAASGYGSVS 1198
1180 SGEIAQNSNDVAVTGGKIYDLKTELENKISSAKTQAQNSLHFSVADQGNFTVSNPYS-1238
1199 TGAFSQASGD-----YGVALGGESEASGAOSTAVG-----AAAGSGDGAFAGALS 1246
1239 -----SYDTSKTSDVITTFAGENGITTKVKNKGVVRVIGIDQTKGLTTPKLTGVNNGK 1289
1247 AEGTESTALGYFASATGESATAVGAESVADGTSAAAFGEGABATSNYST-----1295
1290 GIVIDSQNGQNTITGLSNT-LANVTNDKGSVRTTEOGKIIKDBDKTRAASIVDVLSAGFN 1348
1296 -----ALGYSTASGFNSTALGNFSTAGSSSVAVGGDATASGAYIAAQOASVAS-GVN 1349
1349 LQNGEAV-----DFVSTYDTVNFADGNATTAKVYDTSKTSKVYDVNVDDTT 1398
1350 SVAVGALLGLLPTAESGDFSTALGGAAWAPGLNSTALGNFABSTGESSVALGAD-----1404
1399 IEVKDKKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIDVAHLNLTLSGDIQTAKGAS 1458
1405 -SVADRDFAV-----SVGSAGNERQITNVAAG---TOGTDAV-NLDQLTAVAEATAQGTS 1453
1459 QANSSAGYVDADGNKVIYDSTDNKYKQAKNDGTVDKTEKVAQDKLVAQAQTPDGTQAQWN 1518
1454 KYFKASGSDSDAGAYIEGDNALAAGEGANASSDNSTAVGAGAAQVAENATAVGMDA---1510
1519 VKSVINKEQVNDANKKQGINEDNAFV-----KGLEKAA 1551
1511 LASGTGAAL--GNNAQALGENSSAVGSNALASDGTATANGAQAQAISTYATALGSEAVA 1568
1552 SDNKT-----KNAAVTVGD-----LNAVAQTPLTIFAGD-----1579
1569 SDNQATAVGFRSAAASNVSAAFGYSESSGRLLSALGYCAVASSDYSTAVGAASLASGAS 1628
1580 ---TGTTAKKGETLTIKGGQTDTHKLTNNIIGVAGTGTFTVKLAKDLTN-----LN 1629
1629 AVAVGEFSESTGDESVAVGSGTFFGFIIPARASCTGAAAFGGGAWATADYTTAIGMNSYAD 1688
1630 SYNAGGTKIDDKGVSPVDSS---GQAKANTPVLNANGLDLGGKVISNVGKGTK---DT 1681

```

```

Db      1689  GVNA--SALGQSAALADNALAIGNGRADAIGASVVGVDASATGINSTGVGRQVNVIGE 1746
Qy      1682  DAANVQOLNEVR-----NLLGIG-NAGNDNADGNQVNIADIKKDPNSGSSNRRTVIKAGTV 1736
Db      1747  NAVSVGVNSFVRESAVNGVALGANAGATGADSVAL-----GSGSRT-YEANTV 1793
Qy      1737  LGGKGN-----NDTEKLATGGVQGVVDKGNAGDLSNVVWVKTKDKGSKKAL 1783
Db      1794  SVSGNGRGGPATRIYVNSDGE-VATDAVNVKGQDLAALADVQRTSGMVQTTGGVARAT 1852
Qy      1784  --LATYNAAGTNYLTNNPAEALDRINEQGRFPHVNDGNQEPVV-QQRNGIDSSAGKH 1840
Db      1853  GDRATAAGAT-----ASGARSAVAAGASTASATGASAMGVDSASGVN 1897
Qy      1841  SVATGFOAKADGEAAVIGROT---QAGNQSIAGDQAQATGDSOIAIGTG-----NVV 1891
Db      1898  STAMGRTNSIGENGVALGVNSFVRSQGANAVAGANAGASGADSVAGLSGSRTYEANVV 1957
Qy      1892  A-----GKHSAGIDPST----- 1904
Db      1958  SVSGNGRGGPATRIYVNVGAGATADASTDAINGGQFFQSLSNTASFLGGGAIGAQGVF 2017
Qy      1905  ----- 1904
Db      2018  VAPTYVIQASVYNNVGAALTALDSKVTELDARGGTATATARTVSLRTAAVPAVASATV 2077
Qy      1905  -----VKADNSVSGVNNNQFTDATOTDVFQ----- 1929
Db      2078  ASSVUTSSASVQGTPTAAVVGVTATPAATSTAVGNAVAANHI TGTATGGSAAYAGPNDTA 2137
Qy      1930  VGNNTVITESNVALGNSAISLA-GTHA-----GTQAKKSDGTA-GTTTATAGATGVKGF 1982
Db      2138  IGSNARVNDGSTAVGANTQAAATNAVAMEGAQVSAASGTAIGQARASAOGAVALG 2197
Qy      1983  AQGTAA-VGAVSVGASGAEIRIQNVAAGEVSATSTDAVNGSOLYKATQGTIANATNELDHR 2040
Db      2198  QGSVADRANTVSVSGVSGEORAVNAAG---TRATDAVNVKQL---DSGVAAANSYTDSSR 2251
Qy      2041  I-----HONENKANAG-ISSAMAWAMPQAVIPGRSMVTGGIATHN 2080
Db      2252  YSAMDSEFYQGDIEDRLRQRNRRLDRQAGMGSAMLNMGASVAGIASQNRKIGAGVGFQN 2311
Qy      2081  GQGAVALGSLKSLDNGQVWFKINGSADTQGHVGAAGAGF 2120
Db      2312  GESALSVGYQRAISPRATV--TIGGALSGDDSSIGVGAGF 2349

```

RESULT 15

```

Q8GM75  PRELIMINARY; PRT; 1210 AA.
AC  Q8GM75;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Adhesin.
GN  HIA.
OS  Haemophilus influenzae.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC  Pasteurellaceae; Haemophilus.
CX  NCBI_TaxID=727;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=nonotypable strain 3179B;
RX  MEDLINE=22300350; PubMed=12410830;
RA  Laarmann S., Cutter D., Juehne T., Barenkamp S.J., St Gene J.W.;
RT  "The Haemophilus influenzae Hia autotransporter harbours two adhesive
RT  pockets that reside in the passenger domain and recognize the same
RT  host cell receptor.";
RL  Mol. Microbiol. 46:731-743(2002).
DR  EMBL; AY078089; AAL79954.1; -.
DR  InterPro; IPR008635; HIM.
DR  InterPro; IPR001680; WD40.

```

```

DR  InterPro; IPR005594; YadaA.
DR  Pfam; PF05662; HIM; 4.
DR  Pfam; PF03895; Yada; 1.
DR  PROSITE; PS00678; WD REPEATS 1; 1.
SQ  SEQUENCE 1210 AA; 125073 MW; 6A26A860A947FEA9 CRC64;

Query Match      7.1%; Score 763; DB 2; Length 1210;
Best Local Similarity 24.6%; Pred. No. 1.1e-16;
Matches 356; Conservative 170; Mismatches 451; Indels 470; Gaps 65;

Qy      855  TNGINAGGKA--ITGLSNTLTDA-----NATTGHTVQLGIVDSIDTKTRAA 898
Db      55  TNGINAYGSTNPFADALNNSATDLNRNVPAYSGLLNLNEXTKANKQLLVADST-----AA 110
Qy      899  SIGDVLNAGNLKNGDAKD---FVSTVDTVDFT-NGNATTAKTVYDGK-----ASKVA 948
Db      111  TVGDLRLKLVVSTYKSTKEESNOVKADEVLFTGSGAATVSSKSENGKHTITVSVTKGE 170
Qy      949  YD-VNVVDTTTHLTGADGNKNQIGVKT---TTLTKTDKADKDAKAINFSVNSGDDKALINA 1003
Db      171  FNTVKTTDAATAGAAGA-NTNERGKVTVSGVTDLTKATAE-----DKKQVATV 216
Qy      1004  KDIAADNLNTLAGEIRNTKGTADTALQTF-QVKVKVKGNGDDNDNDADTITVGDKATNQVNT 1062
Db      217  VDVAKAINDAA-----TFVKVENSNEELDNDQQAD-----NADQALKKAGDT 257
Qy      1063  LKLG-KGNGLDIOQNKDG-TVTFGINTQSGLKAGNNTTLNNGLSIKNTAGNEQIOVGAD 1120
Db      258  LTFKAGKN---LKAQRDGRKNIITAL-----ANDLNVTATVSDKLSLGAN 299
Qy      1121  GVKPAKVNVNGVVGAGIDGTTRITRDEIGFA---GTNSGLDKSPHLSKDGINAGGKITN 1177
Db      300  G--NKNV-----VTSITT---DGLNFAKETAVNGD-----TN 325
Qy      1178  IQSGEIAQNSDAVVGSKIYDLKTELENKISSAKTAQNSLHEPSVADEQGNFTVSNPY 1237
Db      326  IQLSGIGSTLTDTITG-----ITKTATN----- 348
Qy      1238  SSYDTSKTSDVITPAGENGITTKVNGKVRVGDIDQTKGLTTPKLTGVNNGNGKIVIDSQN 1297
Db      349  ----- 356
Qy      1298  QONTITGLSNTLANVNDKGSVRTTEQGIKIBEDKTRAASIIVDLVSAGFNLQNGEA-- 1355
Db      357  -----LAASVKDVLVSAGNIRGAATAGG 379
Qy      1356  ---VDFVSTYDTVNFA-DGNATTAKTVYDDTSKTSKVVDVNVVDVTIIEVKOKKLGKVT 1410
Db      380  VDNVDFVSTYDTVDFTDGDGETTITVTYQKANGKGAEV--KIGAKTSVIEKDGKLVTKG 437
Qy      1411  TTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSLGDIQTAKGASQANSAGYVD-- 1468
Db      438  ANKEAGGTNTTP-ATEDMDGKGLVTAETVINAVNNAGWRIKT-----TTANGQAGKETV 492
Qy      1469  ADGNKVIYD-----STDNK--YQAKNDGTV-DKTK-EVAKDKLVQAQTPDGTIAQ 1516
Db      493  TSGTNTVTFDNGNATTAIVSKDDKGNITVKYDVNVGDLKVDATPKIVA-----DTTALT 547
Qy      1517  MNVKSVLNKEQVNDANKKQGINEDNAPVKGLEKAAADSKNTKNAAVTVGDINAVAQTPITF 1576
Db      548  VN-----SGNEANKPKG-----KVAEITNEGDKKKLVNABGLVNALNSLSWT--AT 591
Qy      1577  AGDTGTTAKKLGELTITIKGGQTDITNKLTNNIIGVVAAGTDGFTVKLAKDLTNLSV---- 1631
Db      592  AGKDGTVNAGSADQEVKAGDKVTFKAGD-NLEVKQDGNFTVSLNSTLGLTSLTLDKT 650
Qy      1632  -NAGGTVKIDDKGVSFVDSGQA-----KANTPVLISANGLDGGKVISNVGKGTQDADAANV 1686
Db      651  ANGATTKITKDGTLTTPANGCATGTNNANIISVTTSGISAGNAKAITNVASS----- 701
Qy      1687  QQLNEVRNLLGLGNAGNDNADGNQVNTADIKKDPNSGSSNRRTVIKAGTVLGGKGNNDTE 1746
Db      702  -----LNAYGDT-NTNFDATANSATDLTROFDANGAYNGLJ-----NLNE 740

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:32:39 ; Search time 39.4567 Seconds
(without alignments)
2800.358 Million cell updates/sec

Title: US-09-813-214A-9
Perfect score: 10708
Sequence: 1 MNHIVKIPNKATGTFMAVA.....NGSADTQGHVGAAGAGFHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	630.5	5.9	2249	1 OMPA_RICRI	P15921 rickettsia
2	550.5	5.1	2021	1 OMPA_RICCN	Q52657 rickettsia
3	545	5.1	2660	1 YEEJ_ECO57	Q8x8v7 escherichia
4	544	5.1	2003	1 YDBA_ECOLI	P33666 escherichia
5	508	4.7	2358	1 YEEJ_ECOLI	P76347 escherichia
6	494	4.6	1856	1 OMPB_RICJA	O06653 r outer mem
7	493.5	4.6	1953	1 BIGA_SALTY	P25927 salmonella
8	480.5	4.5	3590	1 FHAB_BORPE	P12255 bordetella
9	480	4.5	1643	1 OMPB_RICPR	Q53020 r outer mem
10	474.5	4.4	1577	1 HLVA_PROMI	P16466 proteus mir
11	466.5	4.4	2432	1 Y43R_IRV6	P18305 chilo iride
12	456.5	4.3	1655	1 OMPB_RICCN	Q9Kka3 r outer mem
13	444.5	4.2	1902	1 P3P_LACLC	P15292 lactococcus
14	437.5	4.1	1902	1 P2P_LACLC	P16271 lactococcus
15	437.5	4.1	1902	1 P2P_LACLC	P15293 lactococcus
16	435.5	4.1	1654	1 OMPB_RICRI	Q53047 r outer mem
17	433.5	4.0	1902	1 P2P_LACPA	Q02470 lactobacill
18	430	4.0	1645	1 OMPB_RICTY	P96389 r outer mem
19	425	4.0	1608	1 HLVA_SERMA	P15320 serratia ma
20	403.5	3.8	1569	1 YPJA_ECOLI	P52143 escherichia
21	398	3.7	1848	1 CBPA_CLOCL	P38058 clostridium
22	380.5	3.6	2334	1 WAPA_BACSU	Q07833 bacillus su
23	366	3.4	1325	1 YDEK_ECOLI	P32051 escherichia
24	360	3.4	1300	1 L20K_RICRI	P14914 rickettsia
25	354	3.3	1176	1 SLAP_BACSH	P38537 bacillus sp
26	353	3.3	1829	1 FRPC_NEIMC	P55127 neisseria m
27	351	3.3	1839	1 ALE3_AZOVI	P44496 azotobacter
28	349.5	3.3	1829	1 FRPC_NEIMB	Q9jyv5 neisseria m
29	342	3.2	1723	1 PM20_CHLPN	Q92812 chlamydia p
30	339	3.2	1025	1 SLAP_CAUCR	P55828 caulobacter
31	338	3.2	1286	1 AIDA_ECOLI	Q03155 escherichia
32	329.5	3.1	1694	1 IGA0_HAEIN	P44969 haemophilus
33	325.5	3.0	1567	1 ICEN_XANCT	P18127 xanthomonas

RESULT 1					
OMPA_RICRI					
ID	OMPA_RICRI	STANDARD;	PRT;	2249 AA.	
AC	P15921;	1990 (Rel. 14, Created)			
DT	01-APR-1990	(Rel. 14, Last sequence update)			
DT	01-APR-1990	(Rel. 14, Last annotation update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmpA).				
DE	antigen) (rOmpA) (rOmpA).				
GN	OMPA.				
OS	Rickettsia rickettsii.				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;				
OC	Rickettsiaceae; Rickettsiae; Rickettsia.				
OX	NCBI_TaxID=783;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=R;				
RX	MEDLINE=90354033; PubMed=2117568;				
RA	Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;				
RT	"A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";				
RL	Infect. Immun. 58:2760-2769(1990).				
CC	- FUNCTION: ELICITS PROTECTIVE IMMUNITY.				
CC	- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.				
CC	- PTM: Glycosylated (Probable).				
CC	- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; M31227; AAA26380.1; -.				
DR	PIR;	A41477;	A41477.		
DR	InterPro;	IPR006315;	Autotransport.		
DR	InterPro;	IPR005546;	Autotransporter.		
DR	Tram;	PF03797;	Autotransporter; 1.		
DR	TIGRFAM;	TIGR01414;	autotrans_bar1; 3.		
KW	Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.				
FT	SIGNAL 1 28 POTENTIAL.				
FT	CHAIN	29	2249		OUTER MEMBRANE PROTEIN A.
FT	DOMAIN	212	1180		13 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	212	286		A (TYPE I).
FT	REPEAT	287	358		B (TYPE II).
FT	REPEAT	359	430		C (TYPE II).
FT	REPEAT	431	505		D (TYPE I).
FT	REPEAT	506	577		E (TYPE II).
FT	REPEAT	578	652		F (TYPE I).
FT	REPEAT	653	724		G (TYPE II).
FT	REPEAT	725	799		H (TYPE I).
FT	REPEAT	800	874		I (TYPE I).
FT	REPEAT	875	949		J (TYPE I).

FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match 5.9%; Score 630.5; DB 1; Length 2249;
Best Local Similarity 23.2%; Pred. No. 5.4e-16;
Matches 500; Conservative 212; Mismatches 779; Indels 665; Gaps 109;

QY 180 LADNLRKRYRTHAQ-----HASTAV-----GAMSYAKG-----HFSNAPGTR- 218
D 1 MANISPLKFKAIQOGLKAALFTTSTAAILMLSSSGALGVATGVIATNNRAFSNVNNGNN 60
QY 219 -----ATAGT-----YSLAVGLTATAKAASSIYAVGNAQAIGFAATAVGSGT 261
D 61 WNETAAGVANGTPAGGPNWAFYGGDYVTADAADRIKAIN-----VAGTT 110
QY 262 QVNIN-----RG-----IALGFGSOVLQKMDVNAANVRAYAPDDNQPIDNRYK 305
D 111 PVGLNITQNTVVGSIITKGNLLPVTNAGKSLTLNGNNAVAANHGFDAPADNY----- 163
QY 306 ATFKNGATVDFSIGNSGNDISIRKIIIVGAGSADTDAAVVAQLEAVRLANROITPKGD 365
D 164 -----TGLNIALGGANAAL-----IIQSAAPSKIITLAGNID-----GGGIITVKTD 205
QY 366 DSNRVEKGLGKTLITIGGAOTSAITDHNIGVQNGDGLKVQLAE-----TLTSLKMTT- 420
D 206 AALNGTTGNTNALATVNVGAGTATLG-----GAVIKATTTKLTNAASVLTLTNANAVLTGA 261
QY 421 -ENLTANEKVTV-----GKTRLTDTKIGFTNDMGIDESKPYLDKDTGHHAGGQKITLTA 475
D 262 IDNTTGGDNVGLNGLALSQVTDIGNTSLATISVG-----AGTATLGGAVIKATT 314
QY 476 GVVDDDAATYQQLKVNQT-----ABSALQTFVKVQKXNDANDSKIIIV-----G 523
D 315 TKLTDAASAVKFTNPVNVVTGAIDNTGNANNGIVTFTGNST-VTGNVGNVTNALATVNVGAG 373
QY 524 KNNKPDG--TOVNTLKLKENGVDVTEFTNGVTFGL-----NQNGLT--VGNSTLANDG 575
D 374 LLQVQGVVKANTINL--TDNASAVFTNPVNVVTGAIDNTGNANNGIVTFTGNSTVTVGD- 430
QY 576 LSVKNTNSKQIOVG-----ADGITFTDISNKPQAGIENTTRI 614
D 431 --IGNTNALATVNVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGA-IDNTT- 485
QY 615 TRDIGFANWTGSLDANKPRLTPTGINAGGKELINVQSAINPATNGQGLDFMNLSTANT 674
D 486 GGDNVGVLNGLALS-----QVTGNIGNTSLATISVGAGTATLGGAV-----IKATTT 534
QY 675 EKSQAATIKOLYNLSQVPLTFAGDTGNVTKLGEILKVKGGKTTADDLTKNN-----IG 730
D 535 KLTDAASAVK--FTNPVVTGAIDNTGNAN--NGIVTFTGNSTVTVGDIGNTSLATIS 588
QY 731 VVADSTNSLTSLVKLAKTLSLDVAVNTKLTASDKV---TVDS-----GNN TAKLO-NGDLTF 783
D 589 VGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNGLALS- 647
QY 784 SKQNTGATPATNS-KTIGVDGLKFTDNNGIALDGTYYITKDKVGFAGQDGLSKPKYLD 842
D 648 --QVTGDIIGNTSLATISV-GAGTATLGGAVIKATTTKLTNAVASVK-----FT 693
QY 843 KDKLVGVEITTINGIN-----AGGKAITG-LSNTLTDTAT-NATTGHVTLQGV---DST 892
D 694 NPVVVTGAIDSTGNANNGIVTFTGNSTVTVGDIGNTNALATVNVGAGTATLGGAVIKATT 753
QY 893 DKTRAASGTDVINAGFNKNNGDAKDFVSTVTDVDFPINGNATTAKVTVD--GKASKVAYDV 951
D 754 KLTNAASVLTLTNANAVLTG--AIDNTTGGDNVGLNGLALSQVTDIGNTSLA-TI 809
QY 952 NVDTGTHLTGADGNKNOIGVKTTLTKTDAKDKKAINFSVNSGDDKALINAKIADNIN 1011
D 810 SVGAGTATLGG-----VIKATTTKLT-----NAASVLTLTNANAV----- 845

QY 1012 TLAGEIRNTKGTADTALQTFQVKVKEKNGD--DDNDADTTITVGKDAKT-----NQVNTLK 1064
D 846 -LTGAVDNTTGGDNVGLNGLALSQVTDIGNTSLATISVGAGTATLGGAVIKATTTK 904
QY 1065 LKGNKGLDIOTNKDGTVTTFGINTOSGLKAGNNT--TLNNGI-----SIKNTAGNEQIOV 1117
D 905 LTNAAASVLTLTNANAVLTGAIDNTTG--GDNVGLNGLALSQVTDIGNTSLATISV 961
QY 1118 GA-----DGVKFAKVNNGVVGAG-IDGTTRITRDEIGFAGTN--GS 1155
D 962 GAGTATLGGAVIKATTTKLTDAASAVRFTNPVVVTGAIDNTGNANNGIVTFTGNSTVGN 1021
QY 1156 LDKSKPHLSKDGINAGGKTIINIOSGEIAQNSNDAVTGGKIYDLKTELENKISSTAKTAQ 1215
D 1022 VGNINA-LATVNVGAG--LLQVQGVVKANTINLTD-----NASAVTFT 1062
QY 1216 NSLHEFSVADEQGNFTVSNPYSDYTSKTSDDVITFAGENGIT-----TKNKGKV 1266
D 1063 NPVVVTGAIDNTGN-----ANNGIVTFTGNSTVTVGNVGNVTNALATVNVG-- 1106
QY 1267 RVGIDQTKG-----LTTPKLTVG-----NNNGKGIIVDSQNGQNTIT 1303
D 1107 -AGLLQVQGVVKANTINLTDNASAVTFTNPVVVTGAIDNTGNANNGIV--TFTGNSTV 1163
QY 1304 G-LSNTLANVTNDKGSVRTEQGGKIIKDEDKTRAASIVD-----VLSAGFNLOQNGEAVD 1357
D 1164 GDIGNTNALATVNVGAGITLQAGGSL-----AANNIDFGARSTLEFNGPLDGGKAIP 1216
QY 1358 F-----VSTYDTVNFADGNATTAKTYTDDTSTKTSKVYV 1390
D 1217 YYPFKGAIANGNAILNVNTKLLTASHLITIGTVABINIGAGNLFTIDASVGD-----VT 1269
QY 1391 DVNVDDTTIEVKDKLGVKTTTLTSTGTGANKPALSNOATGDALVKASDIVA----- 1442
D 1270 ILNAQNINFRARDSVLVLSNL-----TGVGVNILL-----AADUVAPOADEGTV 1314
QY 1443 ----HLNLTSGDIOTAKGASQANSAGYVDADGNK-----VIYDS--TDNKYYQAKNDGT 1491
D 1315 VFNGGVNGLN-----VGSNVAGTARNIGDGGKNKENTLLIYNAVTTITDDVNLGEGINVL 1368
QY 1492 VDKTKFEVAK-----DKLVAQAOTPDGTL--AQMNVKSVINKEQVUNDA 1531
D 1369 INKNADFTSSATFNAGAIQINDATYTTIDANNGLNIPAGNIQFAHAQAQLVLQNSSGNDR 1428
QY 1532 NKQGINEDNAFVKGLEKASDNKTNAAVTVGDLNAV-AQTPLTEAGDTGTT---AKKL 1587
D 1429 TITLGNID-----PDNDEGIVI-----LNSVTAGKKLTIAG--GKTFCGAKHL 1471
QY 1588 GETLTIKG-GQTDNTKLTNNIGVAGTGDGFTVKLAKDLTLNLSVNVAGGTIKDDKGVSVFV 1646
D 1472 -QTILFKGAGDCST-----AGTTENTTIVLDIT----- 1499
QY 1647 DSSQQAANTPVL SANGLDLGGKVISNVGKTKTDDAANVQQLNEVRNLLGLGNAG---N 1703
D 1500 --GC-----LELG-----ATTANVVLFDVAQLTGTGNTGIGGFLD 1531
QY 1704 DNAD-----GNQVNIADIKKDPNSGSSNRVTIKAGTVLGGCKGNNDTEKLA---TGQVQ 1754
D 1532 FNKNGMVTLLNNVNVAGAVQ--NTGGTNNGTLLI-----VLGASNLNRVNGIAMLKVAGN 1585
QY 1755 VGVDKDGNNAN-GDLSNVVWVKTKQDGSKKALLATYNAAGQTNLYLTNNPABADIRNEQIR 1813
D 1586 VTIAGKGVKIGEIQGTGNT-----LTLPAHENLTG-----SINKTGGQALK 1628
QY 1814 PFHVNDGNQBPVVOGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGD 1873
D 1629 LNFVNGSGSVGV-----GTAANSVGDITTAGATSFASBV 1663
QY 1874 NAOAT-----GDOQIA-----IGTGNVVAGKHSIGAIGDPSTVKADNSYSVGNNOFTDATQTD 1926
D 1664 NAKGTATLGGTTSFANTFTNTGATVLAK-----GSITSFAKNVTAT 1704

QY 1927 VFCVGNNTVTSNSVALGNSAISAGTHAGTQAKKSDGTAGTTTGA-----TGTVKG 1981
 Db 1705 SP-VANSATINFSNLAFTSN--ITGG-----GTTTLGANGVYTTGT-GS 1746
 QY 1982 FAGQTAVGAVSVA--SGAERRIQNVAAAGVBSATSTDAVNGSQYKATQGIANATN 2035
 Db 1747 FTDTLTNTLTFDGAAGSGGNILKSGSLDLSGVSTLAL-----VVTATN 1791

RESULT 2
 OMPA_RICCN STANDARD; PRT; 2021 AA.
 ID AC Q52657; P95591; P95593; P95594; Q52667; Q52668; Q52669;
 AC Q52670; Q52674;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
 DE antigen) (rOmpA) (rOmp A).
 GN OMPA OR RC1273.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=94171067; PubMed=8125327;
 RA Croquet-Valdes P.A., Weiss K., Walker D.H.;
 RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
 RT conorii (Malish 7 strain).";
 RL Gene 140:115-119(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RT Science 293:2093-2098(2001).
 RN [3]
 RP SEQUENCE OF 8-204 FROM N.A.
 RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
 RX MEDLINE=97015921; PubMed=882558;
 RA Roux V., Fournier P.-E., Raoult D.;
 RT "Differentiation of spotted fever group rickettsiae by sequencing and
 RT analysis of restriction fragment length polymorphism of PCR-amplified
 RT DNA of the gene encoding the protein rOmpA.";
 RL J. Clin. Microbiol. 34:2058-2065(1996).
 RN [4]
 RP SEQUENCE OF 953-2012 FROM N.A.
 RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
 RA Raoult D., Fournier P.-E., Roux V.;
 RT "Phylogenetic analysis of spotted fever group rickettsiae by study
 RT of the outer surface protein rOmpA.";
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry.
 CC -!- PTM: Glycosylated (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U01028; AAA17405.1; -.
 CC EMBL; AB008674; AAL03811.1; -.

DR EMBL; U43794; AAB49549.1; -
 DR EMBL; U43798; AAB49550.1; -
 DR EMBL; U43806; AAB49551.1; -
 DR EMBL; U45244; AAB49566.1; -
 DR EMBL; U46918; AAB49566.1; -
 DR EMBL; U83440; AAC35176.1; -
 DR EMBL; U83443; AAC35179.1; -
 DR EMBL; U83448; AAC35184.1; -
 DR EMBL; U83453; AAC35189.1; -
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR005546; Autotransporter.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRFAMs; TIGR01414; autotrans_bar1; 1.
 KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
 KW Complete proteome.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 2021 OUTER MEMBRANE PROTEIN A.
 FT DOMAIN 238 946 THR-RICH.
 FT DOMAIN 1424 1528 THR-RICH.
 FT VARIANT 60 60 N -> NN (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 76 76 R -> H (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 86 137 MISSING (IN STRAIN M1).
 FT VARIANT 126 133 MISSING (IN STRAIN MOROCCAN).
 FT VARIANT 953 954 VT -> II (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 1245 1245 D -> A (IN STRAINS INDIAN TICK TYPHUS, M1
 FT AND MOROCCAN).
 FT VARIANT 1308 1308 N -> H (IN STRAIN MOROCCAN).
 FT VARIANT 1877 1877 M -> I (IN STRAIN INDIAN TICK TYPHUS).
 FT CONFLICT 10 10 Q -> K (IN REF. 1).
 FT CONFLICT 92 92 I -> V (IN REF. 1).
 FT CONFLICT 126 126 V -> I (IN REF. 1).
 FT CONFLICT 137 137 T -> N (IN REF. 1).
 FT CONFLICT 157 157 G -> D (IN REF. 1).
 FT CONFLICT 368 369 IS -> VN (IN REF. 1).
 FT CONFLICT 374 388 KATLGGAIIKATTTK -> LLQVGGVVKANTIN (IN
 FT REF. 1).
 FT CONFLICT 640 640 N -> D (IN REF. 1).
 FT CONFLICT 669 669 V -> I (IN REF. 1).
 FT CONFLICT 793 793 N -> D (IN REF. 1).
 FT CONFLICT 803 804 VN -> IS (IN REF. 1).
 FT CONFLICT 809 823 LLRVQGVVKSNTIN -> KATLGGAIIKATTTK (IN
 FT REF. 1).
 FT CONFLICT 898 898 D -> V (IN REF. 1).
 FT CONFLICT 908 908 P -> N (IN REF. 1).
 FT CONFLICT 985 985 N -> K (IN REF. 1).
 FT CONFLICT 1009 1009 L -> S (IN REF. 1).
 FT CONFLICT 1013 1013 Y -> S (IN REF. 1).
 FT CONFLICT 1182 1182 K -> Q (IN REF. 1).
 FT CONFLICT 1314 1314 N -> V (IN REF. 4).
 FT CONFLICT 1451 1451 H -> N (IN REF. 1).
 FT CONFLICT 1624 1624 G -> D (IN REF. 1).
 FT CONFLICT 1628 1628 E -> G (IN REF. 1).
 FT CONFLICT 1872 1872 A -> V (IN REF. 1).
 FT CONFLICT 1875 1875 T -> P (IN REF. 1).
 FT CONFLICT 1878 1879 MS -> LP (IN REF. 1).
 FT CONFLICT 1936 1936 E -> A (IN REF. 1).
 FT CONFLICT 1965 1970 MTAPLP -> ITPPLS (IN REF. 1).
 FT CONFLICT 1997 1997 G -> R (IN REF. 1).
 SQ SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;
 Query Match 5.1%; Score 550.5; DB 1; Length 2021;
 Best Local Similarity 21.5%; Pred. No. 4,1e-13;
 Matches 511; Conservative 243; Mismatches 811; Indels 811; Gaps 116;
 QY 14 GTFMAVYAKHSHSTGGSCATGQGVSVTLTSFARIAALAVLVIGA-----T 60
 Db 88 GDYTTADVADHIITAINVADTTPIG----LNAQNTVGVSVTGGNLLPVITAGKSJT 143
 QY 61 LINGSAYIGISEADGGKGGANARGDKSTAIGDIA-----QALGQSSTIAIGNKIVHNSN 115
 Db 144 LNG-----NNADAANHGFAPADNYTGLGNIALGGANAALIIQSAAPAKITLAGNIN 195
 QY 116 NNANIGAKASGNESIAIGG-DVLASGHASIAIGSDDLLYKKETVQOISELLPIRQOKAL 174

Db 196 GGGIITVKTDAINGTIGNTALATVNVGAGIATLEGAIKATTKLTNAASVL----- 249
Qy 175 NDIYQLADTNLQKRRTHAQGHASTAVGAMSAYAKHESNAPGTR-----ATAGTYSLAVG 230
Db 250 -----TLTNVAVLTGAI DNTTGV DNVGVNLNGALSQVTG 285
Qy 231 LTATAKAASSIAGVSNAAQI GFAATAVGGSTOVNLRGIALGFGSQVLQK-----DNDVNAA 287
Db 286 NIGNTNALATISVGAGKATLGGA--VIKATTKLT DNASAVTFNPNVVTGAIDNTGNA- 342
Qy 288 NVRAYAPDDNQPIDNRKYATPKAGATDVFSIGNSGNDSIRKKIINVAGASAD----- 340
Db 343 -----NNGIVTFTGDSVTGNTGNTNA-----LATISVGAGKATLGGAIIK 383
Qy 341 -----TDVNVAVLKEAVRLA-----NQIITFKGDDSN-----RVEKG 374
Db 384 ATTKLT DNASAVTFNPNVVTGAI DNTGNANNGIVTFTGDSVTGNTGNTNALATISVG 443
Qy 375 LGKTLTITGG---AQTSALTDHNI GVVQNGDGLKVQLAETLTSKMTV-----TENLTAN 426
Db 444 AGKA-TLGGAI I KATTKLT DNASAV-----TFTNPNVVTGAIDNTGNAN-N 488
Qy 427 EKVTVGKTRLT DKIGFTNDMNGIDESKPYLDKDTGHIAG-----GOKITKLTAGVDDDA 482
Db 489 GIVTFTGDSVTGNTGNTNALATI-----SVGAGKATLGGAIIKATTKLT DN 537
Qy 483 ATYQOLKKVQNT-----AESALQFT--VKKYDKNGN DANDSKILTVGKNNKP-DGTQ 532
Db 538 SAVTFTNPNVVTGAI DNTGNANNGIVTFTGDSVTGNTGNTNALATISVGAGKATLGAI 597
Qy 533 VNTLKLXGENGVDVTTETNGVTTFGL-----NONNGLT--VGNSTLNDGSLSVKNTNSNK 585
Db 598 IKATTKLT DNASAVTFNPNVVTGAI DNTGNANNGIVTFTGDSVTGNTGNTNALA 654
Qy 586 QIQVGA-----DG-----ITFTDISNSKPGAGIENTRITRDRDIGGFANN 624
Db 655 TVNVGAGIATLEGAVIKATTKLTNAASVLTNTVNAVLTGA-IDNTTGV--DNVGVNL 711
Qy 625 TGSJD-----ANKPLTFTGNIAGGKELTNVQSAL--NPATNGGQL 663
Db 712 NGALSQVTGNTGNTNALATISVGAGKATLGGAIIKATTKLT DNASAVTFNPNVVTGAI 771
Qy 664 DFMNRLSTANPEKSGSAATIKDLYNLSOVLTFAGDTGPNVTKLGE-----I 711
Db 772 D-----NTGNA-NNGIA-----TFTGDS--TVTGNIGNTALATVNVGAGL 809
Qy 712 LKXVGGKTADDLTKNNI GVVADSTDNLSLTVKLAKTLSLDDAVNTKLTASDKVTVDSG- 770
Db 810 LRVOQ-----GVVKSNTIN-----LTD-----NASAVTFTNPNVVTGAI 843
Qy 771 NNTAKLQNGDLTFGKQN--TGATPATNS-KTIGVDGLKFTDNNGIALDGTYYITKDKVGF 827
Db 844 DNTGNANNGIVTFTGDSVTGNTGNTNALATISVGAGKATLGGAIIKATTKLT DN 900
Qy 828 AKODGSLDKSPYLDKOKLKVGEVEITTINGINAGGKAITGLSNTLTDATNATGHVTQLG 887
Db 901 ----SAVTFNPNV-----VVTGAIDNTGNANNGIV-----TFTGDSVTG 936
Qy 888 IVDSTDKTRAASIGDVLNAGNFKNNG--DAK--DFVSTYDVTDFINGNATTAKVTVDGK 943
Db 937 NIGNTNALATVNVG-----AGVTLQAGGSLDANNIDF--GARSTLEF-NG-----PLDGG 983
Qy 944 ASKVAYDVNDGTTIHLTG--ADGNKQI GVKTTTLTKTDAGDKAKTNFVSNGD----- 996
Db 984 QNAIPY-----YFKGATANGNALLNVNTKLLTAYHLITGTVAEINIGAGNLPAD 1034
Qy 997 ----DKALINAKDIADNLNTLAGBIRNTKGTADTALQTFQVKVKENGDDDN---DADTI 1049
Db 1035 ASAGDVTILNAQDI--HPRAL-----DSALVLSNLTGUVN-----NILLAADLV 1077
Qy 1050 TVGKDAKTNQVNTLKLKKNGLDIOTNKGDTVTFGINTQSGLKAGN--NTTLNNGLSIK 1107

Db 1078 APGVDECTVFPD-----GGVNLNIGSNVAGAA-----RNIGDVGNKNFNTLLIINAVTIT 1128
Qy 1108 NTAGNEQIQVAGDGVKFAKVNNGVVGAGIDGTTTRITREDEIGFAGTNGSLDKSKPHLSKDG 1167
Db 1129 D-----DYNLEGIQNVLINN---NADFTSSAFNAGTQINDATYIIDANNGLN--- 1175
Qy 1168 INAGGKITNIQSGEIAONS--ND-AVTGGKIYDLKTELENKISSTAKTQNSLHESFVA 1224
Db 1176 IPAGNIFAHADAQILIQNSSGNDRTITLGANIDPDNDDEGIVILNSVTA----- 1225
Qy 1225 DEQNNFTVSNPYSSYDTSKTSVITPAGENGITTKYNKGVVRVIGIDOTKGLTTPKLTIVG 1284
Db 1226 ---GKLTIIAGGKTFGGAHLQDIIV-FKGE-----GDFGTAGTF----- 1261
Qy 1285 NNNKGIVIDSQNGQNTITG---LSNTLVNTNDKGSVRTTEQGLIKDEBKTRAAISVD 1341
Db 1262 --NTTNIVLD-----ITQLELGATTANVVLFKDAVOLQTGT----- 1296
Qy 1342 VLSAGFNQNGEAVDFVSTYDVTVNFPADGNATTAKVTVYDDTSKTSKVYVDNVDDTTIEV 1401
Db 1297 ----NIGG-----FLDFNAKNGTIVTLNN-----NVNVAGT----- 1322
Qy 1402 KDKKLGVKTTTLTSTGT---GANKFALSNOATGDALVKASDIIVAHLNTLSGDIQTAKGA 1457
Db 1323 ----VKNTGGTNGTLLI VLGASNL---NRVNGIAMLKVG-----AGNVTIAGK- 1363
Qy 1458 SOANSSAGYVDADGNKVYDSTDNKYQAOKNDGTVDKTEKVAOKLVAQAOTPDGTLAQM 1517
Db 1364 --GNVKIGEIOGTGNTL---TLPAHF--KLITGSINKT-----GGQALK 1401
Qy 1518 NVKSVINKEOVNDANKKQINEDNAFVKGLEKAASDNKTKNAAVTVGDLNNAVOTPLTFA 1577
Db 1402 N-----FMNG-----GSVGVVGTAAANSV 1420
Qy 1578 GDTGTT--AKKLGELTITKG---QDTDNKLTDMNIGVWAGTDGTVVKLAKDLT- 1626
Db 1421 GDITTAGTAFASSVNAKGTATLGGTTSFAHTFTNTGAVTLAKGSIITSFAKNVTATSFVA 1480
Qy 1627 NLNSVNAG---GTKIDDKGV-----SFVDS-----SQOAKANTPVL 1659
Db 1481 NSATINFNSLAFNSNITGSGTTLTLCGANQVYTGTSFTDTLTLNTTDFDGAAGSGGNIL 1540
Qy 1660 SANG--LDLGGKVISNVKGKGTDTDAANVQOLNEVRNLLGNGNAGNADGNQVNIADIK 1717
Db 1541 IKSGSTLDSLQ--VSNLALVVVAT-----NFDMN----- 1567
Qy 1718 KDPNSGSSNRVTVIKAGTVLGGKG---NND-----TEKLATG 1751
Db 1568 ---NISPTKVTYISAETAGLKPENVKITINNDNRVDFDFTDASTLTLPFAEDIAAG 1624
Qy 1752 GVQGVDKDGNANGDLSNV-----WVKTKDGSKK-----ALLATYNAAGQT 1793
Db 1625 ----VIDEDFAPGCPLANIPNAANI KKSLELMEDAPNGSDARQAFNFGMLTLPLEADAT 1680
Qy 1794 NYLTNN---PAAIADRINEQOI-----RFFHVDNGNQEPVQVQNGRIDSSAS 1837
Db 1681 THLMQDVVKPSDITAAVNVNVVVASNISNITARNMDKVQAGNKGPVSSGDEMD----- 1736
Qy 1838 GKHSVAIGFOAKADGEAAVAIGROTQAGNOSIAGDQAQATG-----DQSLAIGTG 1888
Db 1737 -----AKFGAWISPPFVGNATQKMCNIS--GYKSDTTGGTIGTGFDFGVFDDLLVLGLA 1785
Qy 1889 NVVAGK-----HSGAIGDPSTVKAD-----NSYSYGNNNQFTDA----- 1922
Db 1786 YTRADTDIKLKNNTGDKNKVESNIYSILGLYSVPYENLFVEALASYSDNKIRKSRRI 1845
Qy 1923 ---TOTDVFVGNNITVTESNVAL-----GNSAISAGTHAGTQAKKSDGTAGT 1969
Db 1846 ATTLETYGYQTANGKYKSESYTGLMAGYTYMMSENINLTPLAGLRYSTIKDKSYKETGT 1905
Qy 1970 TTTAGATGTVKGFAGQTAGVAVSVGASGAERRIQNVAAGEVSATSTDAVNGSOLYKATQ- 2028
Db 1906 TY---QNLTVKKNYNTFDGLLGAKVS-----SNINVNEIVLT-----PELYAMVDY 1949

QY 2029 GIANATVELDRIHQENKANAGISSAMAWSPQA 2064
 Db 1950 AFKNKVSADARLQ-----GMTAPLPTNSFKQS 1977
 RESULT 3
 YEEV_ECO57
 ID YEEV_ECO57 STANDARD; PRT; 2660 AA.
 AC Q8X8V7; Q8X2B9; Q8X2C0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeev.
 GS 2315 OR EC2775/EC2776.
 ON Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Poracousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533 (2001).
 [2]
 RP SEQUENCE FROM M.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22 (2001).
 CC -!- SIMILARITY: Contains 16 Big-1 domains.
 CC -!- SIMILARITY: Belongs to the intimin/invasin family.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 1315.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE005423; AAG57041.1; -
 CC EMBL; AP002559; BAB36198.1; ALT_FRAME.
 CC EMBL; AP002559; BAB36199.1; ALT_FRAME.
 CC InterPro; IPR003344; Big_1.
 CC InterPro; IPR003535; Intimin.
 CC InterPro; IPR008964; Invasin_intimin.
 CC InterPro; IPR00601; PKO.
 CC Pfam; PF02369; Big_1; 16.
 CC PRINTS; PR01369; INTIMIN.
 CC SMART; SM00634; BID_1; 16.
 CC SMART; SM00089; PKD_8.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834
 FT DOMAIN 840 929
 FT DOMAIN 931 1033
 FT DOMAIN 1042 1132
 FT DOMAIN 1134 1236
 FT DOMAIN 1245 1335
 FT BIG-1 1.
 FT BIG-1 2.
 FT BIG-1 3.
 FT BIG-1 4.
 FT BIG-1 5.
 FT BIG-1 6.
 FT BIG-1 7.
 FT BIG-1 8.
 FT BIG-1 9.
 FT BIG-1 10.
 FT BIG-1 11.
 FT BIG-1 12.
 FT BIG-1 13.
 FT BIG-1 14.
 FT BIG-1 15.
 FT BIG-1 16.
 SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;
 Query Match 5.1%; Score 545; DB 1; Length 2660;
 Best local similarity 20.9%; Pred. No. 9.1e-13;
 Matches 489; Conservative 306; Mismatches 888; Indels 652; Gaps 114;
 QY 178 YQLADTN---LQYRR-----THAQGHASTAVGAM--SYA-KGHFSNAFGRA 219
 Db 412 FDLVDRNNNVILEYRKKEVLRLTLDPVTGKSGEVKSLVSSLOTYALKGY--NVEATAL 469
 QY 220 TABGTVSLAVG----LTATAKAASSIAVGSNAQAIGFAATAVGGSTQVNLNRGIALGFGS 275
 Db 470 EAAGGKVVTGKDILVTLTPAYRFTSTPTDNTWPIEVAEDVKGNFS--NRQSMVVVQAP 528
 QY 276 QVLQKNDVNAANVRAYAPDDNPIDNRYKATFKNGATDVFSIGNSGNDISIRKLIINV 335
 Db 529 TUSQKDSVSLSS-----QTLSDSHSTATLTFTAHDAAGNPVIGLVLSRTH 575
 QY 336 AGSADTDAVNAQLKEAVRLANRQITFKGDDSNRRVEKGLGKTLTITGCAOTSALTTHNI 395
 Db 576 EGVDQ---ITLSDWKD-----NGDGSYTOI-----LTTGAMSGILT---L 609
 QY 396 GVVQNGDGLKVLQIAETLTLKMWTTENLTANEKVTVGKTR-LTTDKIGFTNDMNGIDBSK 454
 Db 610 MPQLNG----VDAKAPAVVNIISVSSRTHSSIKIDKRYLSGNPIEVTVELRD--ENDK 664
 QY 455 PVLKDKDTHAGGQKTKLTAGVDDDAATYQQLKKNQTAESALQTFVKKVQDKNGDA 514
 Db 665 PVKEQKQQLNT-AVSDINVKPGVTTDWKETADGVKYATYATYKSGSLGTLAKLLQNNWED 723
 QY 515 NDSKIITVGKNNKPDGTQVNTLKLKGENGVDDVTETNGTVTF-----GLNNGMLTVGNS 569
 Db 724 LHTAGFLIDAN--PQSAKIATLS--ASNGVLANANAANTVSVNADEGSNPINDHTVTF 780
 QY 570 TLNNDGLSVKVNINSKQIQVGADGITFTDINSKPGAGIENTRITRDG-----IGFA 622
 Db 781 VLSGSATSPNNQNTAK---TDVNGLATFDLKSQKE---DNTVEVTLENGVKQTLIVSFV 834
 QY 623 --NNTGSLIDANKPR-----LTPGTINAGGKELTNVQAINPATNGQGLDFMNR 668
 Db 835 GDSSIAQVDLQSKNEVVADGNDSATMTATVRDAKGNLLNDVKVTFVNSAAKLS----- 890
 QY 669 LSTANTEKSGSAATIKDLIWNLSQVPLTFAGDTGPNVTKLGEILKVKGGKTADDLTKNN 728
 Db 891 QTEVNSHDGIATATLTSKN-----GDY--TVT-----ASVSGSQANQQ 928
 QY 729 IGWVADSTDSNLSLTKLAKTSLDLDVNT-----KTLTASDKVTVDVSGNNTAKLQNGDLTF 784
 Db 929 VIFIGDQSTAAULTSVPS--GDIIVTNPAPLHMTATLQDK-----NGN---PLDKKEITFS 979
 QY 785 KQNTGATPATNSKITGVGDLKFTDNNGIA-----LDGTYITKDKVKVFAKQDGSIDKS 837
 Db 980 VPNDVASRFSISNS-----GKGMTDSNGTAIASLTGTLAGTHMITA-----FLANSVSDT 1030
 QY 838 KP---YLDKOK-----LKVGEVEITNGINAGGKAITGLSNTLTATNATTHVTVQLGIVD 890
 Db 1031 QPMTFVADKDRAVVVLQTSKAEIINGVDE-----TTLTATVKDPPFDNV---VKQLSW- 1081
 QY 891 STDKTPAASIGDVLNAGNLKNGDA-----KDFVSTVDVTD--FINGN-----A 933
 Db 1082 --FRTSPADTQLSLNAR--NTNENGIAEVTLGTVLGVHTAEAILLNGNRDTKIVNIAPDA 1138
 QY 934 TTKVTVYDGKASKVAYDVNVDTGTHLTGADGNKNQOIGVKTTTLTKTDKGDKAINFV- 992

```
Db 1139 SNAQVTLNIPAAQVVTN-NSDSVQLTATVKDPSNHPVAGITVFTMPQ-----DVAANFTLE 1194
Qy 993 NSG-----DDKALINAKDIAINLNTLAGEIRNTKGT-----ADTALQTFQVKV 1036
Db 1195 NNGIAITQANGAEHVTLKGGKAGTHHTVATLGNNASDAQPVTFVADKDSAVVVLQTSKA 1254
Qy 1037 KENGDDNDADTIITVGDAKTNQVNTLKLK-GKGLDDIQ-----TNKQDT--VTF--- 1083
Db 1255 EIIGNGVDETLTATVKDPFDNNAVKDLQVTFSTNPADTQLSQSKNSNDNSVAEVTPEGT 1314
Qy 1084 --GINT-QSGLKAGNNTLLNNGLSIKNTAGNEQIQV-----GADGVKFAKV--- 1127
Db 1315 VLGVHTAEATLTPNGNNDI---KIVNIAPDASNAQVTLNIPAAQVVTNNSDSVQLTATVKD 1371
Qy 1128 --NNGVVAGI-----DGITRIITRDEIGPAGTNGSLDKSPHLSKODINAGGKKIINIQ 1179
Db 1372 PSNHPVAGITVFTMPQDVAAANFTLENNGIAITQAN---GEAHTVLKGGKAGTHHTVATL 1428
Qy 1180 SGE-----IAQNSNDAVTGGKI-----YDLKTELENKISS---TA 1211
Db 1429 SNNTSIDSQPTVFVADKTSALVVLQISKNEITGNGVDSATLTAIVKQDFNEVANLPTVF 1488
Qy 1212 KTAQNSL-----HEFSVADE--QGNNTFVSNPYSSYDTSKTSQ--VITFAGENGIT 1258
Db 1489 STASSGLTLTPGESNTNESGIAQATLAGVAFGEQVTVTASLANNGASDNKTVHFTGDTAAA 1548
Qy 1259 -----TKVNGVVRVGDIDTKGLTPKLTGVNNG--KGIVDSQNGQNTITGLSNTLA 1310
Db 1549 KIELTTPVDSII--AGTPQNSSGSVITATVVDDNNGFPVKVTVNFTSNAAT-----A 1599
Qy 1311 NVINDKGSVRTTEOGKIIEKDKTRAA-----SIVDVLISAGFNLOQNGEAVDFVSTYDTVN 1366
Db 1600 EMING-GQAVTNEQKATVVTNTRSSIESGARDPTVEASLE---NGS-----STLSTSI 1650
Qy 1367 PADGNATPAKYTYDDTSKTSKVVD--VNVDDTT---IEVKDK----- 1404
Db 1651 NVNADASTAHLT-----LLQALFDTSAGDTTNLXIEVKDNYGNGVPQEQEWTLSVSPSE 1704
Qy 1405 -----KLGVKTTTLT-SGTGANKFA--LSNQATGDALVK 1436
Db 1705 GVTFPSNNAIYTTNHDGNFYAFATKAGVYQVATLENGDSMQQOTVTVFVNVANAETSLA 1764
Qy 1437 AS--DIVAH-----LNTLSGDIQTAKGASQANS-----AGVVDADGNKVYDSTDN 1481
Db 1765 ASKDPVIANNDLITLTATVADTEGNAITANSEVFTLPEDVRANFTLGDGCKVVTDT--- 1821
Qy 1482 KYQAKNDGTVDKTEVA-----KOKLVAAQO---TPDGTFLAQM 1518
Db 1822 ---EGKAKVTLKGTKAGAHVTVTASMAGGKSEQLVNVNFIADTLTAQVNLNVTFEDNFIAN-N 1877
Qy 1519 VKSVINKQVNDANKQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNAVAQTPLTAFAG 1578
Db 1878 VGMTRLQATVDNGNPLANEAVTTLPADVSASPTLGGGS-AITDINGKAEVTL----- 1932
Qy 1579 DTGTTAKKLGETLITKG-QQTDNKK-----LTDNNGIIVVAGTDGFTVKLAK 1623
Db 1933 -SGTKSGYPTVTVSNVNVGVSDTKQVLLIADAGTAKLASITSVYSFVVSITTEGAT--MTA 1989
Qy 1624 DLTNINSNAGTKIDDKGVSFVSDSQAKAN-----TPVLISANGLDUGG 1668
Db 1990 SVTDANGNPVEGIKYNFRGTSVTLSSTSVETDDRGFAEILVTSTEVGLKTVSASLADKPT 2049
Qy 1669 KVISNVGKGTDTDAANVQOLNEVNLGLG-----NAGNDNADGNQV----- 1711
Db 2050 EVISRLINAKADINSATTLSLEIPEGQVWVADVAVKAVHNDQFNPILNESVTFSAEPP 2109
Qy 1712 -----NIADIKDPNSGSSNRTVIKAGTVLGGKNNDT-----KIAT 1750
Db 2110 EHMITSQNTVSTDTHGIAEVTMTPERNGS---YMKASLANGSSVEKDLVVIDOKLTLSA 2166
Qy 1751 GGVQGVGDKDGVANGDLSNVWVKTKQKSGKALLATYNAAQGTNYLTNNPAAIDRINEQ 1810
```

```
Db 2167 SSPLIGVNSPTGA-----TLTATILTSANGT-----PVEG-----Q 2196
Qy 1811 GIR-|-----FHVNDGNOEPVQVQGRNGIDSSASGHSHVAIGFOAKADGEAAVA 1857
Db 2197 VINPSTPEGATISGGKVRINSSQAPVV-----LTSNKVGTIVVTASFH----- 2241
Qy 1858 IGRQTQAGNOSIAIGDNAQATGDSIAIGTGNVVVAGKSHGSAIGDPSTVKADNSYSVGNNN 1917
Db 2242 -----NGVTIQTQITIVKVITGNSSTAHVASFADPSTIAATNS-----DL 2280
Qy 1918 QFTDATQTDVFGVGN---NITV---TESNSVALGNSAISAGTHAGTQAKKSDGTAGTTT 1971
Db 2281 STLKATVED--GSGNLLIEGLTVFPALKSGSATILSTLTAFT-----DQNGIATTSV 2328
Qy 1972 TAGATGVTKGFAGOTAVGAVSVG---ASGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQ 2028
Db 2329 RGAITGSVTVSATTAGMQTVDITLVAGPADASQSVLKNRSSLKGDFTDSAEHLVL- 2387
Qy 2029 GIANATNELDHRHONENKANAGISSAMAMAMFQAYIPG-----RSMVTG--- 2074
Db 2388 -----HDISGNPIKVSSEGLEFVQSGTNAFVQVSAIDYSKNFSGEYKATVTGGGE 2437
Qy 2075 GIAT-----HNGQGAVALGSLKSDN-GQWVKING-----SADTQGHVGA 2114
Db 2438 GIATLIPVLNGVHQAGLSTTIQFTRAEDKIMSGTVLVNGANLPTTTPPSQGTGA 2492

RESULT 4
YDBA_ECOLI STANDARD; PRT: 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Nakade S., Nakamura Y., Nishio Y., Oshima T., Saito N.,
RA Saeki G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RC MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Panchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
RT Escherichia coli K-12."
RL Biochimie 73:1361-1374(1991).
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
```


QY 1935 TVTESNSV-----ALGNSAISAGTHAG-----TQAKKSDGTAGTTT 1972
 Db 1504 TVIADGVGTGLIKQSDSINVEGNGNNSSEVHYGDTLDPVPRPNTSVTSGSDEA 1563
 QY 1973 AGA-----TGT-VKGFAGQTAAGVAVS-----GASGABERRIQNVAAGEVSA 2012
 Db 1564 GGSNNLNGVVGTVNGVNSAGKLKVNNASMMGVBEINTGFTAGTADTTVSPDNVVEGS-NL 1622
 QY 2013 TSDAVNGSOLYKATQGIANATNELDRIHON 2044
 Db 1623 TDADAITSTSVVMTAKGSTDASGNDVDTMSKN 1654

RESULT 5
 YEEJ_ECOLI
 ID YEEJ_ECOLI STANDARD; PRT; 2358 AA.
 AC P76347; P94750;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeeJ.
 GN YEEJ OR B1978.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
 RA Takada J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392 (1996).
 CC -1- SIMILARITY: Contains 13 Big-1 domains.
 CC -1- SIMILARITY: Belongs to the intimin/invasin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000289; AAC75042.1; ALT_INIT.
 CC EMBL; D90837; BAA15800.1;
 CC EMBL; D90836; BAA15799.1; ALT_INIT.
 CC EcoGene; EGI3378; yeeJ.
 CC InterPro; IPR003344; Big_1.
 CC InterPro; IPR003535; Intimin.
 CC InterPro; IPR008964; Invasin_intimin.
 CC InterPro; IPR002482; LysM.
 CC InterPro; IPR000601; PKD.
 CC Pfam; PF02369; Big_1; 13.
 CC PRINTS; PR01369; INTIMIN.
 CC SMART; SMO0634; BID_1; 13.

DR SMART; SMO0257; LysM; 1.
 DR SMART; SMO0089; PKD; 6.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834
 FT DOMAIN 840 931
 FT DOMAIN 932 1033
 FT DOMAIN 1042 1137
 FT DOMAIN 1146 1237
 FT DOMAIN 1246 1350
 FT DOMAIN 1351 1448
 FT DOMAIN 1449 1553
 FT DOMAIN 1554 1655
 FT DOMAIN 1661 1754
 FT DOMAIN 1763 1853
 FT DOMAIN 1855 1950
 FT DOMAIN 1952 2053
 FT CONFLICT 105 105 S -> G (IN REF. 2).
 SQ SEQUENCE 2358 AA; 248599 MW; 232249750BPF31ED CRC64;

Query Match 4.7%; Score 508; DB 1; Length 2358;
 Best Local Similarity 20.6%; Pred. No. 1.8e-11;
 Matches 444; Conservative 285; Mismatches 835; Indels 588; Gaps 101;

QY 178 YQLADTN---LQYRR-----THAQGHASTAVGAM--SYA-KGHFSNAFGTRA 219
 Db 412 YDLVDRNNIVLEVRKKELVRLTLTDPVTCGSEKSLVSLQTKYALKG--NVEATAL 469
 QY 220 TAGTYSLAVG-----LTATAKAASSIAGSNAQAIGAFAATAVGSGTQVNLNRGIALGFGS 275
 Db 470 EAAGGKVVTGKDIIVLTPAYRFTSTPETDNTPIEVTAEDVKGNL-SNREQSMVVVQAP 528
 QY 276 QVLQKQNDVNAANVRAYAPDQNPIDNRYKATFGNGATDVFSIGNSGNDSIRKIIINVG 335
 Db 529 TLQKQSSVSLST-----QTLNADSHSTATLTFIAHDAAGNPVGLVLSRTH 575
 QY 336 AGSADTDVAVNAQLKEAVRLANRQITPKGDDSNNRVEKGLKTLITITGGAQTSALTDHNI 395
 Db 576 EGVQD---ITLSDWKD-----NGDGSYQI-----LTTGAMSGTLT---L 609
 QY 396 GVVQNGDGLKVQLAETILSLKMTTENLTANEKVTVGKTR-LITDCKLGFINDMNGIDESK 454
 Db 610 MPQLNG-----VDAAKAPAVVNIISVSSSRTHSSIKIDKRYLGNPIEVTVELRD-ENDK 664
 QY 455 PYLDKDTGIIHAGGQKITKLTAGVDDDAATVQGLKKVNTQAESALQFTVKKVDKNGNDA 514
 Db 665 PVKEQKQQLN-NAVSDINVPFGVTTDKETADGVYKATYATYKSGSLTAKLLMNNED 723
 QY 515 NDSKIITVGNKPKDGTQVNTLKLKGENGVVDVTTETNGTVTF-----GLNQNNGLTVGNS 569
 Db 724 LHATAGFIIDAN--PQSAKIATLS-ASNNGVLANENAAVTSVNVADEGSNPINDHTVTFA 780
 QY 570 TLNNDGLSVKNTSNKQIQVGADGITFDISNKGAGAGIENTTITRDG-----IGFA 622
 Db 781 VLSGSATSFFNQNTAK---TDVNGLATFDLKSKQE---DNTVEVTLNGLVQKPLIVSVFV 834
 QY 623 --NNTGSLDANKRLPTPTGINAGKELTNVQSAINPATNGQGLDFMNLSTANTEKGSQA 680
 Db 835 GDSTAQVDLQKSK---NEVVADGNSVMTATVROAKGNLLNDVMTVFNVSAAEKLQSQ 891
 QY 681 ATKDLVNLISQVPLTFAGDTGPNVTKLGEILKVKGKTKTADDTLTKNIGVADSTNSL 740
 Db 892 TEVNSHDGIATATLTLKNGDYRVT-----ASVSSGSOANQOVNFIQDOSTAAL 940
 QY 741 TVKLAKTSLDLDAVNT---KTLTASDKVTVDSNNNTAKLQNGDLTFSKNTGTATPATNS 796
 Db 941 TLSVPS--GDIIVTNTAPQVMTATLQDK-----NGN-----PLKDKETITFVNDVA-----SK 987
 QY 797 KTIQVGLKFTDNNNGIA---LDGT---TYITKDKVGFPAQDQGLSKSKP---YLDKDK-- 845
 Db 988 FSIISNGKGMVDSNGVAIASLTGTLAGTHMIMARLA---NSNVSDAQPTFVADKDRV 1043
 QY 846 --LKVGEVEITNGINAGGKAITGLSNTLTDAITNATTVHTVTLGIQIVDSTDKTAAASIGDV 903

Query Match 4.6%; Score 494; DB 1; Length 1656;

Best Local Similarity 22.4%; Pred. No. 3.9e-11;

Matches 407; Conservative 196; Mismatches 658; Indels 555; Gaps 96;

```
QY 545 DVTETNGTTFGLNQNNGLTVGNSTLNNGLSVKNTNSKQIOVGADGIFTDISNSKP 604
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  63 NVAVAPNAVIT--ANANNGINLNPAGSFNGLFSNAN-NLAVTVS-----105

QY 605 GAGLNTTRTRDGIIGFANNNTGSLDANKPRITPTGINAG-----GKELTNVQSA-----653
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  106 -----EDTT-----LGFINNAAN-NANRPNLT--LDAGKTLTITGGQITNVQSAATHNA 151

QY 654 --INPATNGGOLDPMWRSLSTANTEKSGSAATIKLYLSQVPLTFAGDTGPNVTKLGEI 711
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  152 QNIVAKFNGGAAIANNDLSGLGTFDGAAS-----TLVFLANPTTKAPLI 199

QY 712 LKVGKGTADDLTKNNIGVVADSTNSLTVKLAKTLDLAVNKTLTASDKVTVDSGN 771
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  200 L-----ADNAL-----IVNGAN 211

QY 772 NTAKLQNGDLTFKONTGATPATNSKTIGVDLKF-----TDNNGIALD--GTYITKD 823
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  212 GTLVNTGTFQVSDKSFATVKAIN---IG-DGQGFMTNATNANLNLQAGGT-----262

QY 824 KVGPAKODGS-----LDKSKPYLDKD-----KLKVEVEITTNAGGKAITGLSNT 871
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  263 -INFNGDTGTGRVLLSKNGAATDFNVTGSLGGNLK-GIIELTVAINGQLIANAGPANA 320

QY 872 LTDATNATTHVTOIGIVSDSTKTRAASI-GEVLNAGPNLKNNGADKDFVSTYTVDFIN 930
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  321 VIGTNN---GAGRAAGFVVSVDNGKAATIDQVY-----AKDMV-----IQSAN 361

QY 931 GNATTAKVYDGGKAVAYDVNVDGTTIHLTGADGNKNQIGVKTTTLTKDAKDKAINF 990
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  362 AN---GQVNF-----RHIVDVGIDGTAFKTA--SIVAITQNS-----NF 397

QY 991 -SVNSGDDKALINAKDIADNLNTLAGETRNKTGTADTALQTFQVKKVKEGDDNDADTI 1049
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  398 GTTDFGNLAQVTPDPTMTLTGNTGDNANPNGTA--GVITFAANGTLASASADANV---452

QY 1050 TVGKDAKTQNTVLKLGKNGLDLQTNKDGTVTFGINTQSLGKAGNNTLNNGLSINKT 1109
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  453 -----AVTNNITATEASGVGVQVLSGTHTABRLG-NAGSVFKLADGTIVNGK-----499

QY 1110 AGNEQIOVGADGVKFAKVNNGVVGAG---IDGTTTRITRDEIGFAGTNGSLDKSKPHLSKO 1166
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  500 -VNGTVLVG-----GVLAAGAITLDSATITGD-INGGGGAALQSIT--LAND 544

QY 1167 G---INAGGKKITNIQSGEIAQNSNDVAVTGGKIYDLKLETKENKISSTAKTAQNSL---HE 1220
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  545 ATKTLTLGGANIISANGGTI---NFQANGGTI-----KLTST---QNNIVVDCD 587

QY 1221 FSAVEQGNFTVSNPNPSSVDTSKTSVITFAGENGITTKVNGKVRVGVIDQTK---GL 1276
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  588 LAIATDQTVVDASS-LTNAQTLTISGTIGIIGANNTTL-----GQFNIQSKTTLNGNV 642

QY 1277 TTPKLTJVGNNN-----GKG-IVIDSQNGQNTITGLSNTLANVTNDKG 1317
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  643 AINELVLGNNGSVQFAHTVLTITFTNAAGGCKLIIPNVVNNTTLAAGTNLGAANPLA 702

QY 1318 SVRTTEQGIKIODEKTRASIVDVLGAGFNLOQNGEAVDFSVSTYDV-----NF 1367
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  703 EINFSGKG-----ARADTVLVN-----GEGVNLATNITTTDANVGSFVFNA 744

QY 1368 ADGNATTAKVYDDTSKTSKVYDVNVDDTIEVKDKKLGKVTTLTSTGTGANKFALS 1427
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  745 GKKNIVSGTGGQGNKFNFTVALD---NGTIV---KFLGNATFNGNTTIAANSTLOISG 797

QY 1428 QATGDALYKA--SDIVAHNLTSLDIQTAQASQANSAGY-VDADGNKVIYDSTDNKYY 1484
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  798 NYTADFIASAGTGIVFVNTGPIVNLNKQAVPNALKQITVSGPQNVV-----NEIG 852
```

```
QY 1485 QAKN-DGTVDKTKKVAKDKLVAQAQTPD-----GTLAQMNKVSINKEQVNDANKKQ 1536
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  853 NAGNYHGAMTDTTIAFENSSSLGAVLFLPSGIPFNDAGNTIPTLIKSTVGNETA-----EG 906

QY 1537 INEDNAFVKGLEKAASDKT---KNAAVTVG---DLNAVAQTPLTTFAGDGTOTTKLGET 1590
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  907 FSPSVTVSGVSDSVIADGOVIGDQNNIVGLGSDNGIIVNATTLTYAG-IGTINNQG-T 964

QY 1591 LTIKGGQTDNKLTDNNIGVVAGTDGF-TVKLAKDLNLASVNAAGTKIDD-----1640
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  965 VTLGGVPTPG-TVYGLGTIGIGASKPKQVTFDTTDDYNNLGNIIATNTTINDGVTVTGGI 1023

QY 1641 --KGVSFVD-----SSQAKANTPVLGANGL-----DLGKVIASNV 1674
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1024 AAGGIAGTDPDKITLGSVNGNANVRPADGIFSNSTSMIVTTKANNGTIVTVLGNFVGN 1083

QY 1675 GKGTQDADAANVQQLNEVRNLLGL-GNAGNDNADGNQVNIADIKKDPNSGSSSRTVKA 1733
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1084 --GDSDFPVASV-RFTGSNNGAGLKNIYSQVIDFGTYNLGIV-----NSNVILG 1130

QY 1734 G--TVLGGK-----GNN-----DTEKLTG-----GVQVGVDKDG 1761
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1131 GSTTAINGKIDLLTNTLTTFAGGTSTWGNNTSIEITTLANGNIGHIVIAEGAQVNTTGG 1190

QY 1762 -----NANGDLSNVVVKTKDKGSK-----A 1782
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1191 TTTINVQDNANFSGTQTVTLIOGGARFNGTLLGPNFTVTGSRFVNYGLIRAAQODYV 1250

QY 1783 LLATYNAAG-QTYNLTNNPAEADIRNEQIGIRPH-VNDGNQEPVQVRNGIDSSASGKH 1840
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1251 ITRTNNAENIVTNDITNSPFGAPGVQNVVTFVFNATNTAAYNNLLAKNSADSANFVGT 1310

QY 1841 SVAIGFOAKADGEAAVAIGTQAGNOSIAIGDNAQATGQSIAGTGNVAGKHSAGIG 1900
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1311 IVTDTSAITNAQLDVAKDIQAQGLNRLGALR-----YLTPEMV-GSEAGAI-1357

QY 1901 DPSTVKA-----DN-SYSGVNNNOFTDATQTFDVGFGVNNITVTSNSVALGSNSAISAGT 1954
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1358 -PAAVAAGDEAVDNVAVGIWAKPFYTDHQSCKGL-----1392

QY 1955 HAGTQAKSKDGTAGTTTAGTATGTVKGFAGTAVGAVSVGASGASGAEIRRIQNVAAGEVSATS 2014
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1393 -AGYKAKTTGTVIGLDTLAN-----NNLMIGA-AIGITKTDIKHQDYKKGD---K 1437

QY 2015 TDAVNG--SOLYKATQIGIAN--ATNELDHRHOMENK-----ANAGISSAMAMAMMPQ 2063
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1438 TD-VNGFSFSLYGAQGFVENFFAGGSAIFSLNQVKNKSQRYFFPDANGMKSQIAAGNYD-1495

QY 2064 AYIPGRSMVTGGIAT-----HNGQGAVAL---GLSKLSDNGQWVFKING-----2104
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1496 -----NMTFGNLTGVYDYNAMQGVLTVPMAGLSYLKSSDE-NYKETGTTVANKQVNSK 1548

QY 2105 -SADTQGHVGAAYCAG 2119
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D  1549 FSDRTDLIVGAKVAGG 1564

RESULT 7
BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; Q9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein biga precursor.
GN BIGA OR STM3478;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028;
```



```
Db 1140 L-----PSSGSDTSTSDTGFETAG-----TLANGTETLNGDVG 1177
Qy 1275 GLTTPKLTGVNNGKVIDSQNGTITGLSNTLANVTNDKGSVRITTEQCKIIKDEBKT 1334
Db 1178 G-----WLYNEAGASLTV-----NGVTIINGANALAN----- 1205
Qy 1335 RAASIVDVLASGFLNQNGEAVDFVSTYDTV-NFADGNATT-----AKVTY-----DDTS 1383
Db 1206 -----YGTILDA-DAISWTSHSLFNEADGSIITDLLTLNGDVTFYNGDFTG 1249
Qy 1384 KTSKVVD-----VNVDDTTIEVKKKL-----GVKTTTLTSTGTGANKFALSNQATGD 1432
Db 1250 SIAGTSYQOEIVNTGDMTVAEDGKSLVSGSFYFNEEDATLTNCSAVE-----GGEINTII 1305
Qy 1433 ALVKASDVAHLNLTLSGDIQTAKGASQANSAGYVDADGNKVIYDSTD----- 1480
Db 1306 NLTRANDSLTOVN-----SGTITATNGYSALTIVNGSNDP-----KWIWNTATGVINGINPDAPL 1360
Qy 1481 ---NKYQAKNDGTVDTKTEYAKDKLVAAQATPDGTTLAQMNVKSVINKEQVN-----DANKK 1534
Db 1361 INLCRGYNFGNGTIN-----VQGDNAVA---ISGTSY--VINLVSGTINVGTEQCKE 1411
Qy 1535 QGINEDNAF-VKLEKASDNKTNAAVTVGDLNVAQTPLTFFAG----- 1578
Db 1412 DGTNGTGLIGIKGNGNATTINNTAD-----GVINVYADDSYAFGGKTKAIINNGEINLLC 1466
Qy 1579 DTGTTAKKLGTLTIKGSQDTNKLTONNI---GVVAGTDGFTVKLAD-----LTN-LN 1629
Db 1467 DSGCDIVAPGTT-----GTQNDHNGTADIVIPDAPTETGSIPTPPADNPAPQOLSIV 1522
Qy 1630 SVNAGGTKIDDKGVSVFVSSGQAKANTPVLSSANGLDLGGKVISNVGKTGDT----- 1681
Db 1523 GTNADG-----SSGTLKANNLVIGDN-----VKYDTGFTSGTADTTVVVDNAF 1565
Qy 1682 DAANVQQLNEVRNLLGLGNA-GNDNADGN-----QVNIADTKDPNSGSS 1725
Db 1566 TGSNIQADNITSTSVVWNAQSQDADGNDVDTWTKNAYADVATDSSVSDVAQALDAGYT 1625
Qy 1726 SNR--TVIKAGTVLGGKNNDEKLTATGCVGVGVKDGNDGNANGDLSNVVWTKQDKGSKAL 1783
Db 1626 NNELYTSLNVT-----TALNSALKQV-----SGAQATTVFREARVLSNPTM 1669
Qy 1784 LATYNAAGQNTYLTNNPFAEADRINEQGRFFHVNDGNQEPVVGQNGRIDSSASGKHSVA 1843
Db 1670 LA--DAAPQ-----IKDG----- 1680
Qy 1844 IGFOAKADGEAAVAIGRQTQ-----AGNQSIAT-----GDNAQATGDSOI 1883
Db 1681 LAFNVVAKGDPRAELGNDTQYDMLALRQTLDTASQNLTLLEYGIARLDGDSKTAGDNL 1740
Qy 1884 AIGTGNVAGKHGAIGDPSVTKADNSYSVGNNGNQFTDATQDVFVGNNITVTESNSVA 1943
Db 1741 TGGYSQFGLKHSWAFBGLANNLSRYDVIN-----LDSSRSVAYGVNKNIDADMKQQY 1796
Qy 1944 LGSNSAISAGTHAGTQAKKSDGTAGTTTTAGATGTVKGFAGQATVAGVSVGASGARRIQ 2003
Db 1797 L-----EPRSEG-AKTFTMWDGALKVTPYAGVKFRHTMEDGYKERSAGDF 1840
Qy 2004 NVAAGEVSATSDAVNSQL-YKATQG-IANATHELDRHONENKANAGISSAMWASM 2061
Db 1841 NLSMNSGNETAVDSIVGLKDYAGKDGWSATATLEGGFNLSYSQSORTASIQGAAG---- 1896
Qy 2062 PQAVIPGRSMVTGGIATHNGQGAVALGSLKSLDNG-----QWVFKINGSAD 2107
Db 1897 -QSFVGDDQKGGV---NGLATIGV---KYSSNDTALHLDAYQW--KEDGISD 1941
```

RESULT 8

FHAB BORPE STANDARD, PRT; 3590 AA.

ID FHAB BORPE

AC P12255;

DT 01-OCT-1989 (Rel. 12, Created)

```
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DB 15-MAR-2004 (Rel. 43, Last annotation update)
GN Filamentous hemagglutinin.
OS FHAB OR BP1879.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
NCBI_TaxID=520;
(1)
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor.";
RL Mol. Microbiol. 4:787-800(1990).
(2)
SEQUENCE FROM N.A.
STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RC MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Leithwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Rabinowitsch E., Rutter S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Sharp S., Simmonds M., Skelton J., Sanders M., Saunders D., Seeger K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
(3)
SEQUENCE OF 1-3261 FROM N.A.
MEDLINE=89202384; PubMed=2539596;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC 1- FUNCTION: Evidence for a role in host-cell binding and infection.
CC 1- SUBCELLULAR LOCATION: Surface.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M60351; AAA22974.1; --
DR EMBL; BX640416; CA642162.1; --
DR PIR; S21010; S21010.
DR InterPro; IPR008619; Fil haemagg.
DR InterPro; IPR008638; Haemagg act.
DR Pfam; PF05594; Fil haemagg; 18.
DR Pfam; PF05860; Haemagg act; 1.
KW Antigen; Hemagglutinin; Complete proteome.
FT CONFLICT 507 508 KQ -> NE (IN REF. 1 AND 3).
FT CONFLICT 1454 1454 A -> P (IN REF. 1 AND 3).
FT CONFLICT 3574 3590 VEDIGKNYRVFYETNK -> SRTSAARTTSSMKPTNR
FT CONFLICT (IN REF. 1).
SQ SEQUENCE 3590 AA; 367519 MW; C00BD8E22C9DB41D CRC64;
Query March 4.5%; Score 480.5; DB 1; Length 3590;
Best Local Similarity 19.9%; Pred. No. 3e-10;
Matches 492; Conservative 306; Mismatches 900; Indels 769; Gaps 110;
```

```
Qy 9 FNKATGTFMAVEYAKSHSTGG-----GSCATGQGVSVRTL-----SFARIAALAVLVIGAT 60
Db 264 YDHATRTATPIACARGAAGAYADGTAGAMYGKHTLVSSDSGLGVLQSLSPSA 323
```


Db 2130 ADMRALGSHLMQKWKDFKAGRGAEIA-----FYPKEQT-VLAAGA 2170
 Qy 1933 NITVTEGNSVALGNSAISAGTHAGTQ--AKKSDGTAGTITTTAGATGTVKGFAGQAVGA 1990
 Db 2171 GLTILS-NGAIHNGENAAQNRPEGLKIGAHSAISVSGSPALRDVGLKELDDIDDLAA 2229
 Qy 1991 VSVG-----ASGAERIRIONVAAGEVSAT-----STDVAVGSGQLYKA 2026
 Db 2230 VLVNPHFTRIGAAQTSGLADGAAGPALARQARQAPETDGMVDARGLGSADALASLASLDA 2289
 Qy 2027 TOGIANATNELDRIHONENKANAGIS--SAWAMASMPQAVIPGRSMVTGGIATHNGGGA 2084
 Db 2290 AQGL-----EVSGR--RNAQVADAGLAGPSAAPAAGAADV-GVEPVTG---DQVDQPV 2338
 Qy 2085 VAVGLSK 2091
 Db 2339 VAVGLEQ 2345

RESULT 9

OMP B RICPR STANDARD; PRT; 1643 AA.
 AC Q53020; Q9ZC00;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB OR SPAP OR SPA OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Breini;
 RX MEDLINE=91045972; PubMed=2122457;
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RA "Characterization of the gene encoding the protective paracrystalline-surface-layer protein of Rickettsia prowazekii: presence of a truncated identical homolog in Rickettsia typhi."
 RT Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241 (1990).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Breini;
 RC Moron C.G., Yu X.J., Walker D.H.;
 RA "Sequence analysis of ompB of Rickettsia prowazekii."
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Madrid E;
 RC MEDLINE=99039499; PubMed=9823893;
 RX Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria."
 RL Nature 396:133-140 (1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN=Breini;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RA "Mapping of monoclonal antibody binding sites on CNBr fragments of the S-layer protein antigens of Rickettsia typhi and Rickettsia prowazekii."
 RT Mol. Immunol. 29:95-105 (1992).
 RL [5]
 RN CLEAVAGE SITE.
 RP MEDLINE=92104668; PubMed=1729180;
 RX Hackett T., Messer R., Cieplak W. Jr., Peacock M.G.;

"Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein of rickettsiae: identification of an avirulent mutant deficient in processing."
 RL Infect. Immun. 60:159-165 (1992).
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M37647; AAA26390.1; ALT INIT.
 DR EMBL; AF161079; AAD42234.1; -
 DR EMBL; AJ235273; CAAL5140.1; -
 DR PIR; D71630; D71630.
 DR InterPro; IPR006315; Autotransport.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRFAMs; TIGR01414; autotrans_barl; 2.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1328 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1329 1643 32 kDa BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREINL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 T TQEAPLTLGA -> INSRSSSYHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 4.5%; Score 480; DB 1; Length 1643;
 Best Local Similarity 22.2%; Pred. No. 1.2e-10;
 Matches 396; Conservative 222; Mismatches 619; Indels 544; Gaps 97;
 Qy 581 TNSNKQIQVGADGITFTDISNSKPGAGIE-----NTRITRDGIGFANNVTSGLDANKPRL 635
 Db 18 TASTATIVAGFGVAM-----GAAMQYNRTTAAATTFDGIQFQAAAG---ANIPVA 66
 Qy 636 TPTGINAGGKELTNVOSAINPAT-----NGQLDFMRLSTANTEKSSAATIKDLY 687
 Db 67 PNSVITA-----NANNPTFTNPNGHLSFLDTANDLAVTINEDT-----TLGIT 113
 Qy 688 NLSQVPLTFAGDTGPNVTYKGLKGVKGGKTTADD-----LTKNNIGVADSTD 737
 Db 114 NIAQAQKFF-----NFTVAAGKILNTGQITVQESASNTINAAQNALTKVHGGAANW 167
 Qy 738 ----NSLITVKLAKTLSDLDVNTK-----TLTASDKVTVDGNNVAKLONGDLTFSKQN 787
 Db 168 LSLGLSGITFAAEPSEVLEFNLTPTQEAPLTLGANSKI-VNNGNGTLNTNGFIQVS-DN 225
 Qy 788 TGATPATNSKTIQVGLKFTDNNNGIALDGTITYITKDKVGFQKQDGLSKSPYLDKDKLK 847
 Db 226 TFA-----GIKTINID-----DCQGLMFSNTP-----DAANTL-----NLQ 256
 Qy 848 VGEVEITTTGAGGKAITGLSN---TLTDATNATTHVHTQLGIVSDTKRAASI---G 901
 Db 257 VGGNTINFNGIDGTGKLVLSKNGAATEFNVTGLGNLKG--GIIELNTAAVAGKLISQ 314
 Qy 902 DVLNAGFNKNN-GDAKDFVSTVYDVFINGNATTAKVYDVGKASVADVNVVG---TTI 958
 Db 315 GAANAVIGTDNGAGRAAGFIVSD-----NGNAATISGQVYAK-NMVIQSANAGGQVTFE 368

QY 1528 VNDANKQGINEDNAFVKGLEKAAADNKTNAAVTVGDLNAVAQTPLTTFAGDTGTTAKKL 1587
 Db 1839 -----SSTFL-----PLAGGTWSGNIITPTGDLISADAPL--VUTSG--ANK- 1877

QY 1588 GETLITKGGQDTNKLNNIIVVAGTGDFTVKLAKDL-----TNLNS--VNAGG---TKI 1638
 Db 1878 -----SYDQSIIANATPSAITGIQ-KIQLAGDLGGSGTTASSPVISSGAILTKM 1928

QY 1639 DD-KGVFVSSGQAKANTPVL-ANGLDLGGKVIS-----NVG----- 1675
 Db 1929 ANLSGNSQIIGSGTSSPNVLTLSGLQISGTVLSNSATLTPPATATTIGGIEMLGD 1988

QY 1676 ---KGTKDADAANVQOLNEVRNLLGLGN-AGNADNAGNQVIA-----DIKDPNSGSS 1725
 Db 1989 LTGSVATAPTIAAGAILTAKMANLSGNSQIIGSSSTTSTPTNLTLGSLQISGTVLSNS 2048

QY 1726 SNRTVIKA-GTVLGG-----KGNNDT-EKLTATGGVQVG--VDRKGNAN--GDLNSVWV 1772
 Db 2049 ATLTVPATATTIGGIEMLGDLTGSVATAPTATGATLTKMANLSGNSQIIGSSSTTST 2108

QY 1773 KTXD-CGSKKALLATYNAGQNTVLTNNPRAIDRINEQGIKPFPHVNDGN--QEPVVOGR 1829
 Db 2109 PTNLTLGSLQISGTVLSNSAT-LTVPATATA---TTGGIEMLGDLTGSVATAPT--- 2161

QY 1830 NGIDSSAGSKHSAVIGFOAKADGAAVAIGRQTQAGNQSIAIGDNAQATGDSIAIGTGN 1889
 Db 2162 -----RAG--AIVLAKWANLSGTSQIIGSSSTTSTPTNLTLGSLQISGTVLSNS 2212

QY 1890 VVAGKHSAGIDPPTVKA-----DNSYSVGNNGNFTDAT-----QTDVFGVGNNT 1935
 Db 2213 LMLLVPSSVNGDLATLNASGOVIDSGVSI--NNSGLTSASLWNAAKLAITNSWFGATN-- 2269

QY 1936 VTENSVALGNSAISAGTHAGTQAK-----KSDGTAGTTTATAGT- 1977
 Db 2270 ---PNTTAPTRDPATSSVLYVGTGDTASLWNGSVYISLIGAKVPTSVTRTFTTSGSGF 2326

QY 1978 ---TVKG-FAGQTVAGVSGVSGAERRIQNVAAGVSATSDA-----VNG 2020
 Db 2327 QISTINGAFVHYSVISITIGVGTSTGVNL---EVSPTSATPASHVING 2375

RESULT 12
 OMPB RICCIN STANDARD; PRT; 1655 AA.
 AC Q9KK3; Q9KK38; Q9XC45;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
 DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
 GN OMPB OR RCI085.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RA MEDLINE=21442074; PubMed=11557893;
 RX Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098 (2001).
 RN [2]
 RP SEQUENCE OF 33-1649 FROM N.A.
 RC STRAIN=Indian tick typhus, and Malish 7;
 RX MEDLINE=20393643; PubMed=10939649;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the

gene coding the outer-membrane protein rOmpB (ompB).";
 Int. J. Syst. Evol. Microbiol. 50:1449-1455 (2000).
 [3]
 SEQUENCE OF 353-1655 FROM N.A.
 RC STRAIN=Malish 7;
 RA Stenos J., Walker D.;
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia australis, the most divergent rickettsia of the spotted fever group.";
 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CL -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
 CC similarity).
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).

 CC EMBL; AE008659; AAL03623.1; -
 CC EMBL; AF123721; AAF34124.1; -
 CC EMBL; AF123726; AAF34129.1; -
 CC EMBL; AF149110; AAD39533.1; -
 CC PIR; E97835; E97835.
 CC InterPro; IPR006315; Autotransport.
 CC InterPro; IPR005546; Autotransporter.
 CC Pfam; PF03797; Autotransporter; 1.
 CC TIGRFAMs; TIGR01414; autotrans_bar1; 2.
 CC Antigen; S-layer; Cell wall; Complete proteome.
 KW CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1335 1655 32 kDa BETA PEPTIDE.
 FT VARIAT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIAT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
 FT CONFLI 353 354 KD -> GH (IN REF. 3).
 FT CONFLI 776 776 F -> S (IN REF. 3).
 FT CONFLI 1159 1159 E -> D (IN REF. 3).
 FT CONFLI 1177 1177 G -> S (IN REF. 3).
 FT CONFLI 1492 1492 H -> R (IN REF. 3).
 SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;
 Query Match 4.3%; Score 456.5; DB 1; Length 1655;
 Best Local Similarity 22.3%; Pred. No. 9.2e-10;
 Matches 388; Conservative 194; Mismatches 663; Indels 494; Gaps 94;
 QY 630 ANKPRLTPTGINAGGKELTNVQSAINPATNGQLDFMNRSLTANRTERKSGSAATIKDL-YN 688
 Db 2 AQKPNFLKKLISAG---LVTASTATIVASFAGSA--MGAAIQQRTTNAVATTVDGVGFD 56
 QY 689 LSQVPLTFAGDTGPNVTKKLGEILKVGKKTADDLTKNNTGVVADSTNLSLVKLAKL 748
 Db 57 QTAVP-----ANVAVPLNAVITAGVNK-----GITLNTPAGSFNGLFLNTA 97
 QY 749 SLDLAVNTKLTATSDKVT--VDSGNNTAKLQNGDLTFPSKQNTGATPATNSKITGVGDKF 806
 Db 98 NNLD-VTVREDTTLGFTTNVNNANHFNLMLNAGKTLTITGGITNVQAAATKNNVVA 156
 QY 807 TDNNGIALDGTYY--ITKDKVGFAPKQDGSGLDKSKPYLDKDKLVGVEIITNGNAGGA 864
 Db 157 QVNNGAIDNNDLQGVGRIDCGAAASTLVFNIAPTTQKAPLILODNAVIVNGANG---- 212

QY 865 ITGLSNTLTADNATTGHVTLQGLVDSITDKTRAASIGDVLNAGFNKNGDAKD--FVST 922
 Db 213 -----TLNVTNGPFI-----KVSSKSPATV-----NVINIGDGGIMFNTD 247
 QY 923 YTVDFPINGNATTAKVTV---DGKASKVAVDVNVDGTTIHLTGA-DGNKKNQIGVKVTTLT 978
 Db 248 ADVNVTNLNQAAGATITFNGTDTGRLVLLSKNAATDENVTGSLGNLKGIIIEPNTVAV 307
 QY 979 KTDAGDKKALINVS---NSGDDKAL-----INAKDIADNLTLAGETR 1018
 Db 308 NGOLKANAGANAAGVITNNGAGRAAGFVSVDMGKVATTIDGQYAKDMVIOQSANAVGQVN 367
 QY 1019 -----NTKGTADTALOTFQVK-KVKENG-----DDDDADATITVKGDAKTNQVNTKL 1065
 Db 368 FRIHVDVGTGDT--TAFKTAASKVATQNSNFGTTDFGNLAAQIIV-----PNTMTL 417
 QY 1066 KGKNGLDIOT--NKOGTVTFGIN--TOSGLKAGNNTTLNN-----GLSIRKTAGNE-- 1113
 Db 418 NGNFTGDASNPGNTAGVITFDANGTLASASADANAVATNNTITAIEASGAGVVQLSGTHAA 477
 QY 1114 QIQVG-----ADG-VKFAKN-----NGVVGAG---IDGTRITRDEIGHGAGTNGSL 1156
 Db 478 ELRLGNAGSVFKLADGTVINGKVNQPTALVGGALAAAGTITLDGSATTIGD-IGNAGGAAL 536
 QY 1157 DKSCKPHLSKDG---INAGGKKITNIQSGEIAQNSNDVAVTGGKIYDLKTELENKISSTAKT 1213
 Db 537 QGIT--LANDATKTLTGGANIIAGNGTI---NFQANGTI-----KLTST--- 578
 QY 1214 AQNSL---HEFSVADQGNNTVSNPYSDYTSKTSVDITFAGENGITTKVKNGVVRVGI 1270
 Db 579 -QNNIVDFDLAIAIDQ---TGVDASSLTNAQT---LTINGKIGTVGANNKTLGQFNI 630
 QY 1271 DQTKGLTTP-----KLTGVNN-----GKG-IVIDSQNGQNTITGL 1305
 Db 631 GSKSVTLSDGDVAINELVINGNAGVQFAHTYLTITTTNAAGGKLIIFNPVNNNTLAT 690
 QY 1306 SNTLANVNDKGSVRTTEOGKLIKDEKTKRAASIVDVLGAGFNLOQNGRAVDPVSYDTV 1365
 Db 691 GTNLGSATNPLAEINPGSKG-----AANVDVTLNVGKGV---NLVATNITTTIDANV 738
 QY 1366 -----NFADGNATTAKVYTDTSKTSKVYDVNVDTTIEVKDKKLGVKT---TTLTST 1416
 Db 739 GSFIFNAGGTNI VSGTVGQOQGNKFNVALD---NGTIV---KPLGNATFNGNTTIAAN 791
 QY 1417 GT---GAN---KFALSNOATGALVKASDIVAHLNLTSGDIQTAKASQANSAGY-VDA 1469
 Db 792 STLQIGGNTADVASADGTG-----IVEFVNTGPTITVTLNKOAFVNALKQITVSG 843
 QY 1470 DGNKVYDSTDNKYYQAKN-DGTVDKTEKVAKDKLVAQAQTP-----DGTLAQMNKVS 1521
 Db 844 FGNVVI-----NEIGNAGNYHGAVDTTIAFENSSLGAVVFLPRGPFNDAGNRIPITIKS 898
 QY 1522 VINKEQVNDANKKQGINEDNAFVKGLEKASDNKT---KNAAVTVG---DLNAVAQTPLT 1575
 Db 899 TVGNKTAT-----GPDVPSVILVGVDSVIADQVIGDQNNIVGLGSDNDIIVNATTL 952
 QY 1576 FAGDTGTTAKLGETLITIKGGQTDI-----NKLTD-NNIGVVAGT- 1614
 Db 953 YAG-IGTINNQG--TVTLSGGIPNTPGTIVYGLGTGIGASKFKQVPTPTDYNLGNIIATN 1010
 QY 1615 ----DGFTVKLA-----KDLTNLSNVNAGGTKIDDKGVSVFDS-----SGOAKN 1655
 Db 1011 ATINDGVTTTGGIAGIGFGDKITLGSVNGG-----NVRFDGILLSHSTSMITTKAN 1064
 QY 1656 TPVLSANGLDLGGKVISNVKGTGKTDDAANVOQLNEVRNLLGLGNAGNADNAGNQVNIAD 1715
 Db 1065 NGTVTV---LGNAFVGNI--GSDTTPVASVR-----FTGSDGGAGLQNGIYS 1106
 QY 1716 IKKD---PNSGSSSRNRTVIKAG-TVLGGKGNNDTEKLATGGVGVGVVDKGNANGDLSNVW 1771
 Db 1107 QVIDFGTYNLGISNSVILGGGTGAINGKINRLTNTLTF-----ASG--TSTW 1152

QY 1772 VKTQKDGSKKALLATYNAAGQNTLYTNTPAEADRINEQGIREFHVDGNEQEPVQVRNG 1831
 Db 1153 -----GNWTSIETTTLTLA-----NGNIGNIV----- 1173
 QY 1832 IDSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQA--TGDQSIATGTGN 1889
 Db 1174 -----ILEGAQVNA-----TTTGTTTIKVQDNANANFSQTQTYTLIQG- 1211
 QY 1890 VVAGKHSIGAIGDPSTV-----KADNSYSVGNNOFTDATQTDVF----- 1928
 Db 1212 --GARENGTLGPFNVYVVTGSRNFRVYGLIRAANQDYVITRTNNAENVVNDIANSFPGA 1269
 QY 1929 -GVGNNT--VTESNSVA-----LGSNSAISAG-----THAGTQAKKSDGTA 1967
 Db 1270 PGVGNVTTTFVNATNTAAVNNLLAKNSANSANFVCAIVTDSAAITNQLDVAK-DIOA 1328
 QY 1968 GTTTTAGA-----TGTVKGFAGQTVAGSVGASGAERRIQNVAAAG-EVSATSTDA--- 2017
 Db 1329 QLGRLGALRYLGTPTETAEMAGPEA-GAIPAAVAGDEAVDNVAYGIWAKPFYVTDHQSK 1387
 QY 2018 VNGSOLYKA--TOGIANATWELDHRIHONENKANAGISSAMAMASMP---QAYIPGRSMVT 2073
 Db 1388 KGLAGYKAKTGWIGLDTL-----ANDNLM---IGAAIGITKTDIKHEDYKKGKTDV 1439
 QY 2074 GG-----IATHNGQGAVALGSLKSLSDNGOMVFEKINGSADTQGHVGAAGVAGFH 2121
 Db 1440 NGFSPSLYCAQQLVKNFFAQSIAIFSLNVQNKSKQRYF-----FDANGNMSQIAAGHY 1493

RESULT 13

P3P_LACLC
 ID P3P_LACLC STANDARD; PRT; 1902 AA.
 AC P15292;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-
 DE associated serine proteinase).
 GN PRTP.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OG Plasmid.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1359;
 RN [1]_LACLC
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
 RC STRAIN=SK11;
 RX MEDLINE=89340435; PubMed=2760036;
 RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
 RT "Primary structure and organization of the gene for a procaryotic,
 RT cell envelope-located serine proteinase";
 RL J. Biol. Chem. 264:13579-13585(1989).
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 CC specificity, although some substrate preference have been noted,
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and
 CC Pro in the P2 position. Best known for its action on caseins,
 CC although it has been shown to hydrolyze hemoglobin and oxidized
 CC insulin B-chain.
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; J04962; AAA03533.1; ALT_SEQ.
 DR HSP; P00782; 2SET.

DR MEROPS; S08.019; --
 DR InterPro; IPR001895; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PSS0847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
 KW Signal; Plasmid.
 FT SIGNAL 1 33
 FT PROPEP 34 187
 FT CHAIN 188 1870
 FT PROPEP 1871 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT SITE 1867 1871
 FT MOD_RBS 1870 1870
 SQ SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;
 Query Match 4.2%; Score 444.5; DB 1; Length 1902;
 Best Local Similarity 20.8%; Pred. No. 3e-09;
 Matches 443; Conservative 219; Mismatches 753; Indels 715; Gaps 100;
 370 RVEGLGKLTITIGGAQTSALTDNHIGVQVQDGLKQV-----LAETILT--SLRWVTEN 422
 3 RKKKGL--SILLAGTVALGALAVLPVGEIQAKAISQOTKSSSLANTVTAATAQAQNTDT 60
 423 LTANEKVTVGKRLTTDKIGFTNDMNGIDSKPYLD----- 458
 61 TAATTNQAIATQLAAKGIDY-NKLNKVVQODIYVDVIVQMSAAPASENGILRTDYSSTA 118
 459 -----KDTGIHAGGQKIITLTVGKNNKPDGTVNTLKLKGBGVDDVT--- 547
 119 EIQQETNKVIAAQAASVKAAVEQVTOQTAG-----ESGYVVGFGFTKRVVVDIPKLKQI 172
 495 AESALQFTTVKVK-----DKNGNDANDSKIIITVGKNNKPDGTVNTLKLKGBGVDDVT--- 547
 173 A--GVKIVTLAKVYPTTDAKANSMAVQAVMSNYKYKGBGVSVSI-----DSGIDPETHK 226
 548 -----TET-----NQTVTFLN--QNN----- 562
 227 MRLSDDKDVKLTKSDVEKFTDTVKHGRYFNSKVPGFNYADNDNTITDDKVDEQGHMVA 286
 563 GLTVGNSLTANDGLSVKNTNSNKOI--OYQADGITTDTISNKPAGAGIENTTITRD-- 617
 287 GIIGANGTGDDPAKSVVGVAPEAQLLAMKVFNSDTSAKTGSATVVSALIEDSAKIGADV 346
 618 GIGFANNTGSLDANKPRILPTGGINAGGKELTNVQSAINPATNGQGLDFMNR----- 668
 347 NMSLGSNSGNQTLDEPAAVQ--NANESGTAAVISAGNSGTSAGTEGVNKKYVGLQDNE 405
 669 -----LSTANTEKSGSAATIKDLYNLSQVPLT-----FAG----- 698
 406 MVGSPGTSRGATTVASAENTDVIITQAVTITDGTGLQLGPETIQLSSHDFTGSPQKKFYI 465
 699 --DTGPNVTK-----KLGEILKVKGGKTTADDLTK-----NNTGVVADSTDNSLT- 741
 466 VKDASGLSKGALADYTADAKGKTAIVRGFSEFDDKQYAAAGAGLIIIVNTDGTATP 525
 742 ---VKLAKTLS--DLDAVNTKLTASDKVTVDSGNN-----TAKLQNGDLTFKQON--T 788
 526 MTSIALTTTFTFGLSSVTGQKLV--DWVTAHPDPSLSGVKITLAMLPMQKYTEDKMSDFT 583
 789 GATPATNSKTIQVGLKFTDNNGTALDGTITTKDKVGFQKQDGL----- 834
 584 SYGPVSN-----LSFKPDI-TAPGGNIWSTQNNNGYTNMSTGSMASPPFAGSQAALLK 634

QY 835 ----DKSKPYLDKDKLVKVG-----EVEITT-----NGINAGGKAITGLSNTILTDATNA 878
 DB 635 QALNNKNPFYAYYKQLKGTALTDFLXTVEMTAQPINLIN-----YNNVIVSPRRQ 686
 QY 879 TTGHTVQLGIVDSTDKTRAASIGDVLNAGFNLNKNG-----DAKDFVSTYDTVDVFINGNAT 934
 DB 687 GAGLVDVKAALDALEKNPSTVVAE-----NGYPAVELKDFSTDKFTKLTFTNRT 736
 QY 935 TAKVY-----DGKA-----SKVAYDVNVVDGTTI----- 958
 DB 737 THELTVMDSNTDITNAVYTSATDPNSGVLYDKKIDGAAIKAGSNITVPAGKTAQIEFTLS 796
 QY 959 -----HLTGADGNKNOI-----GVK-----TT 975
 DB 797 LPKSPDOQFVEGFLNFKGSDGSLNLPYMGFGFDMNDGKIVDSLNGITVSPAGNGFCTV 856
 QY 976 TLTK-----TDAKG-----DKAINFSVNSGDDKALIN----- 1002
 DB 857 PLLKNKNTGTQYVGMVTDADGNKTVDDQAIAF-----SSDKNALYNDISMKYLLRNISNV 913
 QY 1003 AKDIAD-----NLNLTAGEIRNTK-----CTADTAL 1028
 DB 914 QVDILDQGGNKVTVTLSSSTRKKTYYNAHSQQYIYYNAPAWDGYYPQORGNINIKTADGS 973
 QY 1029 QTFQVKKVKGDD-----DNDADTIT-VGDKAKTNQVNT---LKLKKG---NGL 1071
 DB 974 YTRISGVPEGDKQVDFVPFKLDKAPTVRHVALSAKTENGKTQYLLTAAEKDLSGL 1033
 QY 1072 DIQ-----TNKDGTVTFGINTOSGLKAGNNTLLNNGLSIKNAGNEQIQVGADG 1121
 DB 1034 DATKSVKTEINEVNTLDATFTDAGTADGY-----TKIETPLSEQAQA--- 1077
 QY 1122 VKFAKNVGVGAGIDGTTRITRDEIGFAGTNGSLDKSKHLSKDG--INAGG--KKITN 1177
 DB 1078 -----LNG-----DNSAELYLTDNASNATDQASVQKPGSTSFDLIVNGGGIPDKLSS 1126
 QY 1178 IQSGEIAQNSDNAVTTGGKIYDLKTELENKISSYAKTAQNSLHESPSVA--DEGNNFTVSNP 1236
 DB 1127 TTTGYEANTQ-----GGGYTFSGTYPAADVDTYDAQGGKHDLNTTYDAATNSFTASMP 1181
 QY 1237 YSSYDTSKTSDFVITFAGENGIT--TTKVNKGV-----VRVGIQDTKGLT----- 1277
 DB 1182 VTNADYAAQVDLVADKAHTQLLKHFTKVRILMAPTFTDLKFNGSDQTSATIKVGTGTVS 1241
 QY 1278 --TPKLTVGNNGKGIVIDSQNGQNTITGLS---NTLANVTNDKGSVRTTEQKLIKDED 1332
 DB 1242 ADTKVNVGHTVA---ALDAQHHSVDVPVNYGDNITIKVTAIDKDGNTTTEQKLTITSSYD 1298
 QY 1333 KTRAASIVDVLVSAGFNLQNGEAVDFVSTYDT-VNFADG--NATTAKVTYDDTSKTSKVV 1389
 DB 1299 P-----DMLKKSIV-----TFDQGVKFGTNNKFNATSAKF-YDPKGTGIATIT 1337
 QY 1390 YDVNVDDTTIEVKDKKLGVK-----TTTLTSTGTGANKFAL-----SNQATGALVKAS 1438
 DB 1338 GKVKHPTTLTQVDGKQIPIKDDLTFTSLDGLTGKQFPFGVVGWDTTQNKTFQEALSFIL 1397
 QY 1439 DIVAHLNLTSGDIQOTAKGASQANSSAGYVDADGNKVIYDSTDNKYVQAKNDGTVDKTKEV 1498
 DB 1398 DAVA--PTLSLDSST-----DAPVYTDNPFQITGTATDINAQYLS----- 1435
 QY 1499 AKDKLVAQAQTPDGTTLAQMNVKSVINK-EQVNDANKQGG---INEDNAFVKG---LEKAA 1551
 DB 1436 -----LSINGSSVASQVEDININSGPKGPMIAIDQPVKLLGKKNLTVAV 1479
 QY 1552 SDNKTNAAVTVGDLNAVAQTPLTTFAGDTGTAKLGETLTIKGGQDTDNKLTNNIGVV 1611
 DB 1480 TDSBN---TTTKNITVYEPKTLAAPTVPIS-----TTEPAQIVT--LTAN----- 1522
 QY 1612 AGTDGFTVKLAKDLTLNLSVSNAGTKIDD---KGVSFVDSGQAKANTPVLISANGLDLGG 1668
 DB 1523 AAATGETVQVSAD-----GKTYQDVPAAGVT--ITANGTFKSTDLGYGNESPAVD 1572


```
QY 813 A-----LDGTYITKDKVFAKQDGLSKPYLDKOKLKVGEVEIT-----NGINAGGKA 864
Db 646 AYYQKUGTAL-----TDFLKT--VEMTAQIPNDIN----- 675
QY 865 ITGLSNTLTDATNATTGHVTLQGLGIVDSTDKTRAASIGDVLNAGFNKNG-----DAKDFV 920
Db 676 ---YNNVIVSPRQAGLVDAKALDALEKNPSTVVAE-----NGYPAVELKDF 722
QY 921 STYDVTDFINGNATTAKVY-----DGKA-----SKVAYDVNVDTTI----- 958
Db 723 STDKTEKLTFTNSTHETLTQMSNTDNTNAVYTSATDPNSGVLYDKKIDGAAGAAIKAGSNIT 782
QY 959 -----HLTGADG----- 965
Db 783 VPAGKTAQIEFTLSLPKSFQOQFVEGLNFKGSDGSRNLNLYPMGPFDDWNDGKIVDSL 842
QY 966 -----NKNQIGVKTTTLTKTDAG-----DKAINFSVNSGDDKALIN 1002
Db 843 GITVSPAGGNFGTVPLLTNKN-TGTQYVGMVTDADGNQTVDDQAIAP-----SSDKNALYN 898
QY 1003 -----AKDIAD-----NLTLAGETRNK----- 1021
Db 899 DISMKYLLRNISNVQVDILDGQGNKVTTLSSSTNLTKTYNAHSQYIYINAPAMDGT 958
QY 1022 -----GTADTALQTFQVKKVKGDD-----DNDADTIT-VGKDAKTNOVNT 1062
Db 959 YDQRDGNIKTADGSGYTRISGVPEGGDKRQVDFVPFKLDSKAPTVRHVALSAKTENGKT 1018
QY 1063 ---LKLKKG---NGLDIQ-----TNKDGTVTFGINTQSGLKAGNNTLNNGLSI 1106
Db 1019 QYLYLTAERAKODLSGLDAFKSVKTAINEVTNLDAFTDAGTADGY-----TKI 1066
QY 1107 KNTAGNEQIQVAGDGVKFAKNNVGVGAGIDGTRITRIDEIGFAGTNGSLDKSPHLSKD 1166
Db 1067 ETPLSDEQAQA-----LNG-----DNSAELYLTDNASNATQDASVQKPGSTF 1111
QY 1167 G--INAGG--KKITNIQSGELAQNSNDVATGKGYDLKTELENISSFAKTAQNSLHEFS 1222
Db 1112 DLIVNGGIPKISSTTGYEANTQ-----GGGYTFSGTYPAAVDGYTYNAQKGDH 1166
QY 1223 VA-DEQGNNTVSNPYSDYTSKTSVITFAGENG-----TKVKNKG-----VRGI 1270
Db 1167 TTYDAATNSFTASMEVTNADYAAQVDLYADRAHTQLLKHFDTKVRLTAPTFTDLKFNNGS 1226
QY 1271 DOTKGLT-----TPKLTGVNNGKGVIVDSN-----QONTITGLSNTLA 1310
Db 1227 DOTSEATIKVTGVSADTKTVNVDGTV-----ALDAQHHSFVDVPVNYGDNITIK-----V 1277
QY 1311 NVTNDKGSVRTTEQGIKIDEDKTRAASIVDVLGAGFNLCNGEAVDFVSTYDVTNVPADG 1370
Db 1278 TATDEGNT-TTEQKTISSYDP-----DMLK-----NSVTFDQGVTFGANEF--- 1319
QY 1371 NATTAKVYDDTSKTSKVYDVNVDDTTIEVKDKKLGVK-----TTTLTSTGTGANKFAL 1425
Db 1320 NATSAKF-YDPKTIATITGVKHPHTTLTQVDGKQIPKIDDLTFSFTLDGLTGLGKPGFV 1378
QY 1426 -----SNQATGDALVKASDIVAHLNTLSGLDIQTAAGASQANSAGYVDADGNKVIYDST 1479
Db 1379 VVGDTTQNKTFQALFTFLDAVA--PTLSLESST-----DAPVYTNNDENFOITGTAT 1428
QY 1480 DN-KYVQAKNDGT-----VDKTEBAKDKLVAQAQTPDGTGLAQNNAKSVLNKEQVNDAN 1532
Db 1429 DNAQVLSLSINGSSVASQYVDININSKPGHWAIDQ-FVKLEGNKVLTV-----AVTD- 1481
QY 1533 KQGINEONAFVKGLEKASDNKTKRAAVTVGDMLNVAQTPITFAGDGTGTTAKLUGETILT 1592
Db 1482 -----SEDNTTKNTIYVVEPKTL-AAPTV-----TP-----STTEPAKTVTILT 1520
QY 1593 IKGQTDITNKLFDNIGVAGTGTGTVKLAKDLTLNLSNAGTKIDD-----KGVSFVDSS 1649
Db 1521 ANSAAT-----GETVQYSAD-----GGKTYQDVPAAGVT-VTAN 1553
```

```
QY 1650 GQAKANTPVLSANGLDLGGKVI SNVGGTGTDDTAANVQOL-NEVRNLLGLGNAGNDADG 1708
Db 1554 GTEFKSTDLYGNESPAVDYVVTNI-----KADDDPAQLQAAKQBELTNL----- 1596
QY 1709 NQVNIADIKDPNSGSSNRTVIKAGTVLGKGNNDTEKLAGT--GVQVGVDKDGNDAGD 1766
Db 1597 ---IASAKTSLASGKYDDAIT-----TALAAA-----TOKAQATLDQTNASVDSLGTANRD 1644
QY 1767 LSNVWVKTDKDGSKKALLATYNAAGQNTYLTNNPAEAIIDRINEQIGIRFFHFVNDGNQBPVV 1826
Db 1645 LQT-----AINQLAALPADKKTSL-----NQLOSV 1671
QY 1827 QGRNGIDSSASGRHSVAIGFOAKADGEAAVAIGRQTQAGNQSTAIAGNAQATGQOSTAIG 1886
Db 1672 KAALGTDLGNQTPSTGKTFTTAAALDLVA-----QAQAGTQT-----DDHQAT-----LAKV 1719
QY 1887 TGNVAVAGKSGAIGDPSTVKADNSYSVGNNNQFTDATCTDVFVGNNITVTESNSVALGS 1946
Db 1720 LDVAFLAKAEG-----IKAAETAEVGNAK-----DAATGKTWYADIADTLTSGQASADAS 1769
QY 1947 N-----SAISAGTHAGTQAKKSDGTAGTTTTAGATGT---VKGPAGTAVGAV 1991
Db 1770 DKLAHLQALQSLKTKVAAAEAAKTVGKGDTTGTSDKGGQGTTPAPAPGDIGDKGDEG 1829
QY 1992 SVGASGERRIQNVAAEVSATST---DAVNSOLYKATQGIANATWELDHR 2040
Db 1830 SQPSGG-----NIPTNPATTTSTDDTTDRNGQLTSGKALPKTGETTER 1876
```

```
RESULT 15
P2P LACLC
ID P2P LACLC STANDARD; PRT; 1902 AA.
AC P15293;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-
DE associated serine proteinase) (LP151).
GN PRT.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pLP763.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=89313288; PubMed=2501630;
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashina A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
RL from Streptococcus lactis NCDO763.";
RL Mol. Microbiol. 3:359-369(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyze hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14130; CAA32350.1; -.
CC FIR; S06997; S06997.
```

DR HSP; P00782; 2SBT.
 DR MEROPS; S08.019; --
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TRIFRAM; YIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS0947; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
 KW Signal; Plasmid.
 FT SIGNAL 1 33 POTENTIAL.
 FT PROPEP 34 187 POTENTIAL.
 FT CHAIN 188 1870 PII-TYPE PROTEINASE.
 FT PROPEP 1871 1902 REMOVED BY SORTASE (POTENTIAL).
 FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 1867 1871 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1870 1870 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1902 AA; 200139 MW; 488D8844D88CDF7 CRC64;
 Query Match 4.1%; Score 437.5; DB 1; Length 1902;
 Best Local Similarity 21.2%; Pred. No. 5.4e-09;
 Matches 457; Conservative 201; Mismatches 728; Indels 773; Gaps 105;
 Db 370 RVEGLGKLTITGGAQTSALTDHNGVQNGDLKQV-----LAETLT--SLKMTVTEN 422
 3 RKKGU--SILLAGTVALGALAVLPGVEIQAKAITSQTKGSSLANVTVAATAKQAATDT 60
 423 LPANEKVTYVGTRLTDDKIGFTNDMNGIDESPKYLD----- 458
 61 TAATTNQAIA-IQLAAKGIDY-NKLNKVOQDIYDVIVQMSAAPASENGTLRTDYSSTA 118
 459 -----KDTGHAGGQKITKLTAGVDDDAATYG-----QLKKNQOT 494
 119 EIQQETNKVIAQAASVKAAVEQTQOTAG-----ESYGVVNGFSTKRVVVDIPKLKQI 172
 495 ASSALQTFVKKV-----DANGDANDSKLITVGKNNKPDGTQVNTLKLKGENGVDT-- 547
 173 A-GVKTVLAKVYPTDAKASMANVQAVWSNYKKGEGTVSVI-----DSGIDPTHKD 226
 548 -----TET-----NGTVTFGLN-QNNGLTVGNSTLND-GLSVK 579
 227 MRLSDDKVKLTKSDVEKFTDTAKHGRYFNKVPYGFNVADNNDITDDTVDEQGHMVA 286
 580 NTNSNKQIQVGADG-----ITFTDISNKPQA-----GIENNT 612
 287 GI-----IGANGTGDDPAKSVGVGAPEAQLAMKVFTNSDTSATTSATLVSAIEDSA 339
 613 RITRD--GIGFANNTGSLDANKPRLTPTGNGAGBELTNVQSAINPATNGGQLDFNR-- 668
 340 KIGADVILNLSGSDSGNQTLDEPELAQV--NANESGTAAVISAGNSGTSAGTEGVNKDY 398
 669 -----LSTANTEKSGSAATIKDLYNISQVPLT-----FAG-- 698
 399 YGLQDNEMVGTPTCSRGAATVASENTDVIQTAVITDGTGLQGPETITQLSSNDFGSGF 458
 699 -----DTGPNVTK-----KLGEILKVGKGTATDDLTK-----N 727
 459 DQKFFVVKDASGNLSKGVADYADAKGIAIVKRGELTFADKQYQAAGAAGLIIVN 518
 728 NLGV-----VADSTNSLTVKLAKTL----- 748
 519 NQGTATPVTSMAITTTTPTFGSSVTGQKLVDWAAHPDDSLGVKIALTLVFNQKYTEDK 578
 749 -SDLDVAVN-TKULTASDKVTVDSGNNTAKLQNGDLTFSKQNTGATP-ATNSKTIQVDGLK 805
 Db 579 MSDFTSYGPVSNLSFKPDITAPGNIWSTQNNGNYTNMSGTSMASPFIAQSQALLKQALN 638
 806 FTDNNGIA-----LDGTYITTKVKVPAKQDGLSKPYLDKDKLVKGEVIT--NG 857
 639 NKNPNFYAYIKQLKGTAL-----TDFLKT--VEMNTAQFIND 673
 858 INAGGKAITGLSNTLTDAFNATTGHVTQLGIVDSTDKTRAASIGDVLNAGFNLKNG-- 914
 674 IN-----YNNVIVSPRQAGLVVDVKAIDALEKNEFSTVVAE-----NGYPA 715
 915 -DAKDFVSTYDVFINGNATTAKTY-----DGKA-----SKVAYDVNVDTTI 958
 716 VELKDFSTDKTKFLTFTNRTTHELTYQMSDNTDNNAVTSATDPNSGVLYDKKIDGAAL 775
 959 -----HLTGADG----- 965
 776 KAGSNITVPAGKTAQIEFTLSLPKSPDQOQFVEGFLNFKGSDGSRNLNLPYMGFFGQDNDG 835
 966 -----NKNQIGVKVTKTTLTKTDAGK-----DKALNFSVNSG 995
 836 KIVDSLNGITYSPAGNFGTVPLLTNNK-TGTYQYGGMTDADGNTVDDQALAF--SS 891
 996 DDKALIN-----AKDIAD-----NLNTLAGEIRNKK----- 1021
 892 DKNALYNDISMKVYLLRNLSNVQVDILDGQGNKVTLSSTNRKKTYYNAHSQQYIYYHA 951
 1022 -----GTADTALQTFQVKKVKGENDD-----DNADIT-VGKDA 1055
 952 PAWDGTYDQDRGNKTADDGSYTYRISGVEGGDKRQVDFVFPFKLDSKAPTRHVVALSA 1011
 1056 KTNQVNT--LKLKGGK--NGLDIQ-----TNKDGTVTFGINTQSLGKAGNNTTL 1099
 1012 KTENGKTQYLLTAEKADSLGLDASKVKTAINEVNLDATFTDAGTTADGY----- 1063
 1100 NNGLSINKNTAGNEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFAGTNGSLDKS 1159
 1064 ---TKIETPLSDEQAQA-----LGNG-----DNSAELYLTDNASNATDQDASVQ 1104
 1160 KPILSKDG--INAGG--KKITNIQSGEIAQNSNDATVGKIIYDLKTELENKISSTAKTAQ 1215
 1105 KPGSTSFOLIVNGGGIPDKISSTTTTGYEANTQ-----GGGTYTFSGTYPAADVDTYDAQ 1159
 1216 NSLHFSVA-DROGNNFTSNPYSSYDTSKTSDDVITFAGENGI---TTKVNKGV----- 1265
 1160 GKHLNLTYYDAATNSFTASMPVNIADYAAQVDLYADKAHTQLLKHFDTKVRLTAPTFTD 1219
 1266 --VRVGIDQTKGLT-----TPKLTGVNNGNGKGIVIDSQN-----GONTIT 1303
 1220 LKFNNGSDQTSKATIKVTGTVSADTKTVNVGDTVA--ALDAQHHFSDVPVNVYDNTIK 1276
 1304 GLSNTLANVNDKGSVRITTEQKIIKDEDKTRAASIVDVLSAGFNLQNGEAVDFVSTVD 1363
 1277 -----VTATDDEGNT-TTEQKTIITSSYDP-----DMLK-----NSVTFDQGVTFG 1315
 1364 TVNEADGNATTAKVYDDTSKTSKVYDVNVVDTTIEVKDKKLGK-----TTTLTSTCT 1418
 1316 ANEF--NATSAKF-YDPKGTGATTTGKVKHHTTILQVQDKQIPIKDDITFTGLDLGLTL 1371
 1419 GANKFAL-----SNQATGDALVKASDIIVAHNLTLSDGIQTAKGASQANSAGYVDADGN 1472
 1372 GQKPGVVVVDGTTQNKTFQALTFILDAVA--PTLSLSDST-----DAPVYTNPNPF 1421
 1473 KVIYDSTN-KYYQAKNDGT-----VDKTEKAVDKLVQAQOTPDGTLAQNKKSVINK 1525
 1422 QITGATDNAQYLSLSINGSSVASQYVDININSGKPGHMAIDQ-PVKLLEGKXNLTV-- 1477
 1526 EOVNDANKKQGINEONAFVKGLEKAASDNKTKNAATVVDGLNVAQOTPLTFAGDTCTTAK 1585
 1478 -AVTD-----SEDNTTKNTITVYEPKKTLL-AAPT-----TP-----STTEP 1513
 1586 KLGETLTIKGGQTDNKLTDNNIGVAGTGTGFKLAKDLTNLNSVNAAGTKIDD--KG 1642
 1514 AKTVILTANSAT-----GETVQVSAD-----GGKTYQDVPAAG 1547

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:33:49 ; Search time 70.2522 Seconds
(without alignments)
2905.511 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 MNHIYKVFINKATGTMAVA.....NGSADTQGHVGAAGAGPHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1246.5	11.6	2059	2 D82671	surface protein XP
2	986	9.2	1588	2 A86036	probable adhesin Z
3	986	9.2	1588	2 H91188	probable adhesin E
4	728	6.8	1190	2 A82615	surface protein XP
5	670	6.3	1107	2 AC0976	probable autotrans
6	669	6.2	3705	2 AD0123	probable autotrans
7	651.5	6.1	2340	2 B71704	cell surface antigen
8	645.5	6.0	5291	2 F90696	hypothetical prote
9	634	5.9	5188	2 B85547	probable RTX fami
10	630.5	5.9	2249	2 A41477	190K surface antig
11	622	5.8	6713	2 B89921	hypothetical prote
12	598	5.6	3013	2 AB0480	probable invasin Y
13	590.5	5.5	2271	2 F90073	hypothetical prote
14	568.5	5.3	2554	2 AB3528	hypothetical prote
15	550.5	5.1	2021	2 A97859	extracellular seri
16	549.5	5.1	2535	2 AC0304	190-KDa cell surfa
17	548	5.1	4919	2 T31105	probable hemolysin
18	546.5	5.1	2020	2 C48399	hypothetical prote
19	541	5.1	2660	2 E85822	ABC-type transport
20	539.5	5.0	2703	2 H81193	probable invasin Z
21	536.5	5.0	4152	2 T31102	hemagglutinin/hemo
22	534	5.0	3624	2 AD0835	filamentous hemagg
23	533	5.0	658	2 AH0110	large repetitive p
24	530	4.9	3890	2 C89921	probable surface p
25	529.5	4.9	1910	2 AF0394	hypothetical prote
26	527.5	4.9	5627	2 C83339	probable adhesin h
27	521	4.9	3295	2 AB0074	hypothetical prote
28	519.5	4.9	2468	2 AB3412	hypothetical prote
29	517.5	4.8	4936	2 AH2515	hypothetical prote

30	517	4.8	2481	2 D90011	FmtB protein [impo
31	517	4.8	3029	2 S76109	hypothetical prote
32	513	4.8	2514	2 F81045	hemagglutinin/hemo
33	508	4.7	1536	2 A43855	high-molecular-wei
34	508	4.7	2383	2 D64962	probable membrane
35	500.5	4.7	1651	2 JC1340	outer membrane pro
36	489.5	4.6	1477	2 B43855	high-molecular-wei
37	487	4.5	2479	2 F87386	conserved hypochet
38	483	4.5	1999	2 AB2018	hypothetical prote
39	480	4.5	1643	2 D71630	outer membrane pro
40	478.5	4.5	3591	1 S21010	filamentous hemagg
41	477.5	4.5	2055	2 T31110	extracellular matr
42	477.5	4.5	3535	2 B83641	probable hemagglut
43	476	4.4	2893	2 A64556	toxin-like outer m
44	474.5	4.4	1577	2 A35140	hemolysin A precu
45	467	4.4	1487	2 AG2560	hypothetical prote

ALIGNMENTS

RESULT 1

D82671

surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: D82671

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequer

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; PMID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: D82671

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2059 <SIM>

A;Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN001

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Pacincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

M.; Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF1529

Query Match

Best Local Similarity 11.6%; Score 1246.5; DB 2; Length 2059;

Matches 578; Conservative 344; Mismatches 780; Indels 681; Gaps 108;

QY	62	NGSAYAGIGISADGG-----KGGANARGDKSIAT-----GDIAQA	97
DB	36	NAQVINDGM---DGGCORIVNGSAGVETVATQCSDEPMTVTGYARFFGSPGTAAEQ	92
QY	98	LG-SQSIAIGNKIVHNSNNNANIKAKAGNESIAIG-GDVL-----ASGHASIAIGSDDL	151
DB	93	QGASRNLTJGGSGLYV---NSQGVGVNDVLNKYSIRNGSVITMTVAGTNAIGS---	145
QY	152	YLKKTVOQIISLPIIRGQXALNDIYQLADTNLQYRTHAQGHASTAVGA--MSYAKG	209
DB	146	-----AQSSAADALKASLTKASG	164
QY	210	HFSNAGFTRATAGTYSIAVGLTATA-KAASSIAVGSNAQAIGFAATVAGGSGTQVNLNG	268

165 ARAIAICAKASADGVDTVLGSAGTAGTGASSIAIGLASAVN-GAVAVGGGALVTVPDG 223 Db
269 -IALGFGSOVLKQDNVNAANRAYAPDDNQPIDNRYKATFKGATDVFGISGNXGDSI 327 Qy
224 AVALGLNSVA-----STGKGLSGYDP-KTKTTSTDASAAWKSTLAAV-SIGDVSSTNLK 275 Db
328 RKLIINVAGSADTDVNVVQALKEAVPLNR--QITEFGDDSNRVE-----KGLKCTL 379 Qy
276 TROLGLAAGTSNTDAVVAQLKVVDIEASRGWMLTASGANGNVAPGSSVDLKNK 335 Db
380 TIT-----GCAQTSALTDHNGVQVQGD-----GLKVQLAET---LTSLKMYTTE 421 Qy
336 TITKAIGSNDVQFNLNKDVKTTLAVGDALLNTDGIAGLTDVSLTSTGLAITDGPVATAS 395 Db
422 NLTRANEKV-----TVGKTRLTDTKIG-----GLKVQLAET---LTSLKMYTTE 421 Qy
396 GIDAGSKVISHVAAVAVSET--STDAVNGSOLNAVQVQASQPVFTTNEGAVKRESLQSV 453 Db
454 -----KPYLDKDTG-IHA-----GGOKITKLTAGVVDDD 481 Qy
454 VISGESSTAGTYSGGNLKSVVDEAAGRIHLQADSPKPGNVVINGGKISGVITAGTEED 513 Db
482 AATYGOLKKNQTAESALQFTTVKVDKNGNDANDSKIITVGKKNKPDGTQVNTLKLGE 541 Qy
514 AVNFSQLKSI STAVD---QGWTLTASGNG-----SKVASGG-----TVDLKNT 554 Db
542 NGVDVTTETNG---TVTFGLNQN---NGLTVGNSTLNDGLSVKNTSNKQIOVGADGIT 595 Qy
555 DG-NLTIKSKSGSDSDVFNLSKDFKVDGMTSGTTVANNNGVKV-----GSDVALGTTGLT 608 Db
596 FTDISNKPAGIENITRITRDGIGFANNTSGSLDANKPRLTPTGINAGKELTNVQSAI- 654 Qy
609 ITD-----GPAVTASGIDAGSKVISHVAAV 635 Db
655 ----NPATNGQLDFMNLSTANTEKSGSAATIKDLNLSQV-----PLTFAGDGPVNTK 706 Qy
636 SETSTDAVNGSQL-----NAVQVQASQPVFTTGNEGA-VKR 670 Db
707 KLGEILKVGKKTADDLTKNIGVADSTNLSLTVKLAKT-----QNGDLTFKQNTGA 747 Qy
671 SLGQSVVISGSSSTAGTYSGGNLKSVVDEAAGRIHLQADSPKPGNVVINGGKISGVTA 730 Db
748 -LSPLDANVTKL---TASDK---VTVDSGMNTAKL-----QNGDLTFKQNTGA 790 Qy
731 GTEETDAVNFSQLKSI STAVDQGWTLTASGNGSKVASGGVTLKNTDGNLTISKSGDGN 790 Db
791 TPATN-SKTIGVDGLKFTDNNGLADGTYITKD--KVGFAKQDGLSKPYLDKXKL 847 Qy
791 DWFNLSKDFKVDGM-----TSGTTVVNNDGKVG-----SDVALGTTGLT 831 Db
848 VGE-VEITNGINAGGKAITG-----LSNTLTDATNATTHGVTOLGIYDSTDKTFAASIG 901 Qy
832 ITDGPVAVTASGIDAGSKVISHVAAVAVSETSTDAVNG-----SOLNAVQVQASQPVFTT 886 Db
902 DVLNAGFNLKNGDAKQFVSTVDTVDFTNGNATTAKVTDGKASKAVAYDNNVDGTTIHLT 961 Qy
887 ---NEGAVKRESLQSV-----VISGSSSTAG-TYSGNLKSVVDEAAG--RIHLQ 930 Db
962 GADGNKNOIGVKTTLTKTDKDKKAI NFNVSNGSDDKALINAKO IADNMLNTLAGIRN-T 1020 Qy
931 LADSPK--FG-----NVVING-----GKISGVT 952 Db
1021 KGTADT-ALQTFQVKKVXKNGDDNDADTIITVGDKAKTNOVNTLKLKNGLDIOTNKG 1079 Qy
953 AGTEETDAVNFSQLKSI-----STAVDQGWTLTASGANGSKVASG--G 993 Db
1080 TVTFTGINTQSLKAGNNTLNNGLSIIKNTAGNEQIOVGADG--VKFAKVVNNVGVGAGID 1137 Qy
994 TV-----DLKNTDGNLTISKSGSDSDVFNLSKDFKVDGMTS 1030 Db
1138 GTTRITRDEIGFA-----GTNGSLDKSPHLSKDGINAGGKIKTNIQSGEIAQNSDAVT 1192 Qy
1031 GTTVVNNNDGVKVGSDVALGTTGLTITANGPAVTASGIDAGSKVISHVAAVAVSETSTDAVN 1090 Db

1193 GGIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNFTVSNPYSSYDTSKTSIDVITFA 1252 Qy
1091 GSOLNAVQVQASQPVFTG-----NEGAVKRESLQSVVIS-----GESSTACTYS 1135 Db
1253 GENGITFTKVNKGVRVIGIDQTKGLTTPKL--TVGNNGKGIIVIDSQNGQNTITGLSNYLA 1310 Qy
1136 GGN-LKSVVDAAAGRIHLQLA--DSPKFGNVVINGGK-----ISGVT----- 1175 Db
1311 NVTNDKSGSVTTTQGGKIIKDDEKTRAASIVDVL SAGFNLOQNGEAVDFVSTYDTVNF-A 1368 Qy
1176 -----AGTBETDAV---NFSQLKSI STAVDQGWTLTASGANGSKVASGGVTLKNT 1223 Db
1369 DGNATTAKVTYDDTSKTSKVVYDVNVND-----DTTIEVKDKKLGVKTTTITSTGTGA 1420 Qy
1224 DGNLTISK-----SGSDNDVFNLSKDFKVDGMTSGTTVVNND--GVKVGSDVALGTTG 1275 Db
1421 NKFALSNOATGDALVKAASDIVAHLNLTLSGDIQTAKGASQANSAGYVDADGNKVIYDSTD 1480 Qy
1276 LTIANGPAVTASGIDAGSKVISHV-----AAGAVSETSTDAVNGSOLNAVQVQASQ 1326 Db
1481 NKYYQAKNDGTVDKTEKVAKDLVAQACTPDGTLAQMNKSVINKEQ----- 1527 Qy
1327 PVTFTG-NEGAVKRES--LGOSVVISGESSTAGTYSGGNLKSVVDEAAGRIHLQADSPK 1383 Db
1528 ---VNDANKKQGI---NEDNAP---VKLEKKAASDNKTKNAAVTG-----DL 1566 Qy
1384 GNVVINGGKISGVITAGTEETDAVNFSQLKSI STAVDQGWTLTASGANGSKVASGGVTL 1443 Db
1567 -NAVAQPLTFAGTGTAKKLGELTIK-----GGQTDINKLTNNIGVAVG 1613 Qy
1444 KNTDGNLTISKSGSDNDVFNLSKDFKVEITAG-----NTVVNTDGVKSGVSDVSLGAMGLFANGPSV 1676 Db
1614 TDGF-----TVKLAKDLNLNLSVNAAGTKIDDKGVSFVDSGQAKANTPVL SANGLDLGGK 1669 Qy
1504 HSSTSSVKTLANGESVVRNRTVVNGDGNVND--VVVNDLGLSIVGASLTLGSLINAGSH 1561 Db
1670 VISNVGKTDTDAANVQQLNEVENLJLGL--NAGNDNADGNQV---NIADIKK-DPN--- 1721 Qy
1562 KITNTVAGTETDVAVNFSQLKSVSEAVDKGWTLTASGANGSKVSGGTVDLKNTDGNLAI 1621 Db
1722 SGSSNRTVI-----KAGTVLGGKGNNDTEKATGGVQVGVQDKDGNANG-----DLSNV 1770 Qy
1622 SKSGSDNDVFNLSKDFKVDVETAG-----NTVVNTDGVKSGVSDVSLGAMGLFANGPSV 1676 Db
1771 WVKTKDGSKKALLATYNAAGTNYLTNNPA---EADIRINEQGRFPHVNDGNOEPVVQ 1827 Qy
1677 TASGFNAGDK---VISHVAVGMADTDVAVNSQLQAVQSVTVKATRYYSTNDGG---TQ 1729 Db
1828 GRNGIDSSASGHSVAIGFQAKADCEAAVAIGRQTA--GNQSIAGIDNAQATGDSOIAIG 1886 Qy
1730 GGNVDGIGATGSKAIAAGVGTQSGEGAAAVGSGAAASGKSTAI GRNIAIASADGSVALG 1789 Db
1887 TG-----NVVAKHSG-----AIGDPSTVKADNSYSVGNQNTQDATT----- 1923 Qy
1790 DGAKDGGGGAESYTGKYSQVQNTVGTVSGVDAKGETRESISNVADEKAMDVNLRLD 1849 Db
1924 -----QT-----DVFQV--GNNTVTES---NSVALGNSAISAGTHAGTQ 1959 Qy
1850 AVAKSNLQTDMDRHEINNIEDVFKITGDSASSVKMGVNAALGTNAAVS-GTESVAL 1908 Db
1960 AKKSDGTAGTTTTAGATGTVKGFAGQTAAGVAVSVAAGSABERRIQNVAAGEVATSTDAVN 2019 Qy
1909 GKNTVNSADNAVAIG-NGSVADRA-----NSVSVSGSGSERQVTVNVAAG---TABTDVAVN 1959 Db
2020 GSOLYKATQGTANATNELDRIHQENKANAGISAMAMAWASMPQAYIPCRSMVTGIIAT 2079 Qy
1960 VSOL---NQGLITAKQYTDGMVGNLRRRTSGGVAIAIATANLPQAVOGRGMTSVGVSSY 2016 Db
2080 NGQAVAVGLSKLSIDNGOWFVKINGSADTQGHVGAAGVAGPHF 2122 Qy
2017 QGQSAIAGVSNVAVSESHVFKFSGSANTRSHVGVGAGVGYQW 2059 Db

RESULT 2

A86036
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: A86036
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: A86036
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1588 <STO>
A/Cross-references: GB:AE0051174; NID:gl2518349; PIDN:AAG58749.1; GSPDB:GN00145; UWGP:Z5029
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: Z5029

Query Match 9.2%; Score 986; DB 2; Length 1588;
Best Local Similarity 22.9%; Pred. No. 5.3e-31;
Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;

QY 1 MNHIYKVFNKATGTFMAVAYAKGSHSTGGSCATGQVGSVRTLSFARIAALAVLVIGAT 60
DB 1 MNKIFKVINWPATCNVTVTSETAKSRGKSGSK-----LLISALVAGGM 45
QY 61 LNGSAVAGIGISEADGGK---GANARGDKSIAIGDIAQALGQSQTGAIDGNKIVHNSNN 117
DB 46 L--SSGALANAGNDGQGVVDYSGSAGDQWVAIGKAKA-----NTFMNTSGSS 93
QY 118 ANTG--AKASGNEISIAIGPVLASGHASIAIGSDDLYLKKTQQISELLPIIRGQKALN 175
DB 94 TAVGYDALAEQVSSAIGSKTHAIGGSMAFG-----VSAISE----- 131
QY 176 DIYQLADTNLQKYRTHAQGHASTAVGMSYAKGHFSNAGFTRATAGTYSVLAVGLTATA 235
DB 132 -----GDRSIALGASSYSLGQYSMALGRYSKALGKLSIAMGSSKA 172
QY 236 KAASSTAVGNAQAIGFATAVCGSTQVNLNRGIALGFGQVLOKQNDVNAANVRAYAPD 295
DB 173 EGANAIALGNATKATEIMSGALGDITANASKAYSMALG-ASSVASEENAI-AIGAETEAEE 230
QY 296 DNQPIDNRYKATPKNG-----ATDVFSIGNSGNDSIRRKIINVAGSADTDVAVNA 347
DB 231 NATAIGNNAKAKGTNSWAMGFGSLAKQNTIALNGSQALADNAIAIGQ-----NKA 283
QY 348 QLKEAVRLANRQITFRKGDSSNNRVERGLKTLTITGGAQTSALTDHNIGVVQNGDGLKQV 407
DB 284 DGVDAIALGN-----GSQS-----RGLN--TIALGTASNATGDKLSALGSSS----- 324
QY 408 LAETLSLKMVTENTLNEKVVTVGKTRLTDDKIGTNDNMGIDESKPYLDKDTG--IHA 465
DB 325 -ANGINSVALGADSIADLONTVSGNSSL---ERKIVNVKNKAIGKSDSY-DAINGSQLYA 379
QY 466 GGQKITKLITAG--VWDDDAATYQLKKVQNTAESALQTETVKKVDKNGDAN---DSKI 519
DB 380 ISDSVAKRIGGGGAARDVDDGTV-----TAPTNLKNGSKNNVGAALAVLDENT 427
QY 520 I-----TVGKNNKPDGTQVNTLKLKNGGVDTVTGTGTGTVTFLGNQNNGLTVGNSTLNNDG 575
DB 428 LQMDQTKGKYSAAHGTSSTPASV-----ITDVADGTIS----- 460
QY 576 LSVKNTNSNKQIQVGADGITFTDISNPKGAGIENTRITRDGIGFANNGTSLDANKPRL 635
DB 461 ASSKDAVNGSQLKATNDNDE-----ANTANIATNTSNTATNTATNTATNTNI 507
QY 636 TPTGINAGGKELTNVQSAINPATNGGOLDPFMNRILSTANTEKSGSAATIKDLYNLSQVPLT 695
DB 508 T-----NLDTOSVGDLQ-ADALLMNETKAFSA----- 534

Qy	696	PAGDTGNTVKLGEILKLVKGGKTTADDLTKNNIGVVDSTDSNLSLTVKLAKTLSDLD	755
Db	535	-----HGQDTTSK-----ITNVKDA-----DLT-----ADSTD-AVNGSQLKTINDAVATN	574
Qy	756	TKTLTASDXKVTVDSGNNNTAKLQNGDLTFSKQNTGATPATNSKTIQVUGLKFDTNNGI--A	813
Db	575	TTNI-----ANNISNIANTNTNISMLTETVT-----NUGEDALKWDKONGVFTA	618
Qy	814	LDGTTYITDKVGFPAKGDGSLDK-SKPYLBDKDLKVGEVEITTINGINAGGKAITGLSNTL	872
Db	619	AHGTG--TTSKIINVK--DGLTITGSTDAVNGSQLKTINDAVATNTTNIATNT--TNISLNT	674
Qy	873	TDATN-----ATTGHVTLQIGIVDSTDKTRAASIGDVLNAG--FNLKXNG	914
Db	675	ETVTNLGEDALKWDKONGVFTAAHGNNNTASKITNILDGTVTATSSDAINGSQLYDLSSN-	733
Qy	915	DAKDFVSTYDVFINGNATTAKVYDYGKASKVAYDVNVDTGTHLTLTGADGNKNOIGVKT	974
Db	734	-----IATY-----FGNAS-----VNTDGVFTGPTYKIGETNYNVN--	765
Qy	975	TTLTTKTDAKGD--KAIN--FVSNSGDDKALINAKDIADNLTILAGEIRNTKGTADTALQT	1030
Db	766	-----GDALAAINSSFSLSIGD--ALL-----WDTAGKFSAXHGT-----	799
Qy	1031	FOVKKXKENG-----DDNDADTIYTGKDAKTQVNTLKLKGRN-----GLDIOQNKD	1078
Db	800	-----NGDASVITVDADGEI--SDSSSDAVNGSQLHGVSYYVVDALGGCAEVNAD	847
Qy	1079	GTVTFGINTOSGLKAGNNTTLNNG--LSIKNTAGNEQIOVGADGVKFAKVNNGVVGAGI	1136
Db	848	GTITAPTPTTIA-----NAYDINVGDALNAIDITLDDALLWDAD-----	885
Qy	1137	DGTTTRITRDEIGPAGTNGSLDKPHLSKDGINAGGKKITNIOSGEIAQNSNDAVTGCKI	1196
Db	886	-----AGENGAF--SAAH-----GKDKTASVITNVANGALISAASSDAINGSQL	926
Qy	1197	YDLKTELENKISSTAKTIAQN--SLHEFSVADBGQNNFTVSNPYSSYDTSKTSVITPAG	1253
Db	927	YTTNKYIADALGDPAEVNADGTTIPTAPTITIANAEYNN-----VGDAKDALD	972
Qy	1254	ENGITTKVNGVVRVIGIDQTKGLTTPKLTGVNNGKGIVIDSQNGQNTITGLSNTLANVT	1313
Db	973	DNALLW-----DETA-----NGGAGAYNASHDGKASI-----ITNVA	1004
Qy	1314	NDRGSRVTRTBQKIIIKDEDKTRAASIVDLSAGFNLQNGEAVDFVSTYDTVNFADGNAT	1373
Db	1005	N--GSI-----SEDS-----DAVNGSQLNAT	1024
Qy	1374	TAKVTYDTSKTSKVVDV--NVDDTTIE-----VKDKLGVKTTTITSTGTGANKF	1423
Db	1025	NMMI-----BQNTQIINQLAGNTDATYIQENGAGINVRVNTDGLAFENDASAOQVGATAI	1079
Qy	1424	ALNQATGDALV-----KASDIVAHLNTLSGDIOT--AKGSAQANSAGSYVDADGNKVI	1475
Db	1080	GYNSVAKGSSVAIGOGSYSDVDTGTALGSSSVSSRVIAKGRSDTS-----ITENGVVIG	1134
Qy	1476	YDSTDNKYQAKNDGTVDKTKVEAKDKLVAQAQTDPGTLAQMNVKSVINKEQVNDANKKQ	1535
Db	1135	YDITDCELLGALSIG-----DDGKYRQ	1156
Qy	1536	GINEDNAFVKLEBKAASDNKTKNAAVTVGDL-----NAVAGTPTTF--AGDTGTTAKKLGE	1589
Db	1157	IIN-----VADGSEAHDAVTVROLQWAIQAVATPTPKYFHANSTEEDSLAVGT	1204
Qy	1590	TLTIKGGQDTPNKLTDNNTGVVAGT-----DGFVTKLAKDLTNLNSVNAGTKKIDDKG	1642
Db	1205	DSLAMGAKTIVN--GDKGIGIGYGAYVDANALNGIAGISNAQVITHVNSTAINGSGSTTTRG	1262
Qy	1643	VSF-----VDSGQAKANTPVLASGLDLGGKVISNVGKGTDKDDTAANVOQLNEVRNLL	1696
Db	1263	AQNTYTYAYMDAPQNSVGFVSUSAD-----GQRQITNVAGSADTDVAVNVQOL-----	1311
Qy	1697	GLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVVIKAGTVLGGKGNNDTEKATGGVQVG	1756

Db	1312	-----KVTDAQVSQNTQSITN-----LDNRVTNLDSEVT--NIENG	1345	Db	173	EGANAIALGNATKATEIMSIATGDTANASKAYSALG-ASSVASEENAI-AIGAETEAAE	230
Qy	1757	VDKGNANGDLNVVVTQKDGSKKALLATYNAAQTNYLTNNPABAIIDRINEQIRFFH	1816	Qy	296	DNQPIDNRYKATFKNG-----ATDVFSIGNSGNDSIRRKIIINVAGSADTDVAVVA	347
Db	1346	I-----GDIV-----TTGSTKYFKTN-----	1361	Db	231	NATAIGNAKAKGTNSMAMGFGSLADKVNTIALGNQSALADNAIAIGQG-----NKA	283
Qy	1817	VNDGNQEPVQGRNGIDSSAGKHSVAIGFOAKADGEAAVAIGRQTQAGNQSIAIGNAQ	1876	Qy	348	QLEKAEVRLANRQITFKGDDSNRVEKGLGKLTLTITGGAQTSALTQDHNIQVGVQNGDGLKVQ	407
Db	1362	-----TDGVDPASAQKDSVAI-----GSGSIAAADN--	1387	Db	284	DGVDATIALGN-----GSGS-----RGLN-----TIALGTASNATGDKSLAGSNSS-----	324
Qy	1877	ATGDQSIAIGTGNVAGKHSIGAIDPSTVKADNSYSVGNNNQFTDATQTDVFGVGNNTV	1936	Qy	408	LAETLTSLKMMVTENLTANBKVVVGKTRLTDTDKIGFTNDMNGIDESKPYLDKDTG--IHA	465
Db	1388	-----SVALGTGSV-----ATEENTISVGSSTNQ	1411	Db	325	-ANGINSVALGADSIADLDNTVSVGNSSL--KrkIVNVKNGAIKDSY-DAINGSOLYA	379
Qy	1937	TESNSVALGNSAISAGTHAGTQAK-KSDGTAGTCTTTAGATGTVKGFAGQTAVCASVGA	1995	Qy	466	GGQKITKLTAG--VVDDDAATYQOLKKVQNTAESALQTFVKKVYDKNGNDAN-----DSKI	519
Db	1412	RRITNVAAGRNA-----TDVNVAAQLKSSSEAGGVRYDTKADGSD--YSNITLG-----GG	1460	Db	380	ISDSVAKRLGGGAADVDDGTV-----TAPYNLKNGSKNVGAALAVLDENT	427
Qy	1996	SGAERRITNVAAGREVSATSTDAVNGSOLYKATQGIANATN-----ELDRHIONENKANAG	2051	Qy	520	I-----TVGKNNKPDGTQVNTLKLKKGENVDTTETNGTTFVFLGNQNNGLTVGNSTLNDG	575
Db	1461	NGGTITRISNVAG--VNNDVNVYAQLKQSVQETKQYTDQRMVEMDNKLSKTESKLSGG	1517	Db	428	LQWDQTKGKYSAHGTSSPTASV-----ITDVADGTIS-----	460
Qy	2052	ISSAMAMASMPQAYIPGRSMVTGGIATHNGGAVAGLSKLSDRGQWVFKINGSADTQGH	2111	Qy	576	LSVKNTNSNKQIQVAGDGIITFTDISNKPAGAGIENTTITRDGIGFANNWTSGLDANKPRL	635
Db	1518	IASAMAMTGLPQAVTPGASMASIGCGTIVGESAVALGVSMVANGRWYKLGQSTNSQGE	1577	Db	461	ASSKDAVNGSQLKATNDVE-----ANTANIATNTSNIATNTANIATNTTNI	507
Qy	2112	VGAAGVAGGFH 2122		Qy	636	TPTGINAGGKELTNVQSAINPATNGGQOLDPMFNRLSTANTEKSGSAAITIKOLYNLSQVPLT	695
Db	1578	YSAALGAGIQW 1588		Db	508	T-----NLTDVSGDLQ--ADALLMNETKKAFSA--	534
RESULT 3				Qy	696	FAGDTGPNVTKLGEILKVKGKTTADDLTKNNIGVVADSTDSNLSLTVKLAKTLSDLDVN	755
H91188				Db	535	-----HGQDTTSK-----ITNVKDA-----DLT-----ADSTD-AVNGSQLKTTNDVATN	574
probable adhesin ECs4480 [similarity] - Escherichia coli (strain O157:H7, substrain R1MD				Qy	756	TKLTATASDKVTVDSGNNVAKLQNGDLTFSKQNTGATPATNSKITGVGLKPTDNGI--A	813
C;Species: Escherichia coli				Db	575	TTNI-----ANNTSNIATNTNISNLTEVT-----NLGEDALKWDKONGVFTA	618
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001				Qy	814	LDGTYITKDKVGFAPAKGDSLDK-SKPYLDKDKLVGEVEITTINGINAGGKAITGLSNTL	872
C;Accession: H91188				Db	619	AHGTE--TTSKITNVK-DGDLTTGTSTDAVNGSQLKTTNDVATNTANIATNT--TNISLT	674
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.				Qy	873	TDATN-----ATTGHVTQLGIVDSTDKTAAASICDVINAG--FNLKNG	914
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.				Db	675	ETVNLGEDALKWDKONGVFTAAGNNNTASKITNILDGTVTATSSDAINGSOLYDLSN--	733
DNA Res. 8, 11-22, 2001				Qy	915	DAKDFVSTYDTPFINNATKAVTYDGKASKVAYVNVVDGTTTTLTGADGNKKNQIGVKT	974
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc				Db	734	-----IATY-----FGGNAS-----VNTDGVFTGPTYKIGETNYVNV--	765
A;Reference number: A99629; MUID:21156231; PMID:11258796				Qy	975	TTLTKTDAKD--KAIN--BSVNSGDDKALINAKDIADNLNTLAGEIRNTKGTADTALQT	1030
A;Accession: H91188				Db	766	-----GDALAAINSFSFSLGD--ALL-----WDATAGKFSAKHGT-----	799
A;Status: preliminary				Qy	1031	FOVKKXKENG-----DDNDADTITVGKDAKTNOVNTLKLKGN-----GLDITQTKD	1078
A;Molecule type: DNA				Db	800	-----NGDASVITDVADGEI-----SDSSSDAVNGSQLHGVSSVYVDALGGGAENVAD	847
A;Residues: 1-1588 <HAY>				Qy	1079	GTVTFGINTQSGLKAGNNTLINNG--LSIKNTAGNEQIQVGADGVKFAKVNNGVVGAGI	1136
A;Cross-references: GB:BA000007; PIDN:BAR37903.1; PID:g13363955; GSPDB:GN00154				Db	848	GTITAPTPTTIA-----NADYDNVGDALNAIDTTLDDALLWDAD-----	885
A;Experimental source: strain O157:H7, substrain R1MD 0509952				Qy	1137	DGTTITRDEIGFAGTNGSLDKSKPHLSKDGINAGGKKTNIQSGEIAQNSNDVATGSKI	1196
C;Genetics:				Db	886	-----AGENGAF--SAAH-----GKQKTA SVITNVANGALISAASDAINGSOL	926
A;Gene: ECs4480				Qy	1197	YDLKTELENKISSAKTAQN--SLHBFVSVADEQNNFTVSNPYSSYDTSKTSBIVITPAG	1253
Query Match				Db	927	YTTNKYIADALGGDAENVADGTITAPTPTTIANAEVNN-----VGDAIDALD	972
Best Local Similarity				Qy	1254	ENGITTKVNGVVRVGIDQTKGLTTPKLTGVNNNGKGIIVDSQNGQNTITGLSNTLANVT	1313
Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;							
Qy	1	MNHYIKYIFNKATCTFMVAEYAKSHSTGGSCATGQVGSVRTLSFARIAALAVLIGAT	60	Db			
Db	1	MNKIFRWMPATGNYVTSETAKSRGKKSRSK-----LLISALVAGM	45	Qy	61	LNGSAYAGIGISEADGGKG---GANARGDKSIAIGDIAQALGQSIAIGDNKIVHNSNN	117
Qy	46	L--SSFGALANAGNDGQVDYSGSGADGWVAIGKGAKA-----NTFMWTSGSS	93	Db			
Qy	118	ANIG--AKASGNESIAIGDVLASGHASIAIGSDLLYLKKTETVQOISELLPIIRGQKALN	175	Db	94	TAVGYDAIAEGQYSAIGSKTHAIGGASMAFG-----VSAISE-----	131
Qy	176	DIYQLADTNLQKRYRTHAQGHASTAVGAMS YAKGHFSNAPGTRATAEGTYS LAVGLTATA	235	Db	132	-----GDRSIALGASSYSLSQYNSWALGRYSKALGKLSIAMGDSKSA	172
Qy	236	KAASSIAVGSNAQAIGFAATAVGGSTQVNLNRGIALGFGSOVLQKDNVDNAANRAYAPD	295				

Db 509 NDVT---FNLATALEA-DSLTTGNTAMTTDGVTVGSNVT---LSTGLVITD----- 553
Qy 1400 EVKDKKLVKVTTLTSTGCTGANKFALSNQATGDA---LVKASDIVAHNTLS-CDIQTA 1455
Db 554 -----GPSVSSGSIAGNOKITNVAAGTADTDAVNPQIQAVSSSTASKGNILLAS 603
Qy 1456 GASQAN---SSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKVAKDKLVA--CAQT 1509
Db 604 GANSNVVPGESVDLKNSDGNLLTKTDS-----NDVTFNLATALKVSLTTGNTAMT 657
Qy 1510 PDGTLAQMNKVSINKEQVNDANK--KQGINEDNAFVKGLEKAASDNKTNAVAATVGDNLN 1567
Db 658 TDGVTVGSNVTLGSTGLVITDGPVSTSGI-----SAGNOKITNVAAGTADT 705
Qy 1568 AVAQTPLTFAGDTGTTAKKL-----GEFLITKGGQTDNKLTDNNIGV--VAGTDGTFVKL 1621
Db 706 AVNFSQIQAVSSTASKGNILLASGANSNVAPGESVDLKNTDGNIVISKESGSDVLFNL 765
Qy 1622 AKDLTNLSNVNAGGTTKIDDKGVSF-----VDSSGOAKANTPVLKANGLDLGGKVISNVGK 1676
Db 766 SSSL-KDKLTVGDTVMTGTVGSGVTLGSMGLVITDGPVSTSSGINAGSOKITNVA 824
Qy 1677 GTKDDTAANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRTVVIKAGTV 1736
Db 825 GTADTDAVNLSQLN-----GGNYNGD-----TA 840
Qy 1737 LGGKGNNDTEKLTAT--GGVQGVVDKNGNANGDLSNVVVKTKQDGSKALLATYNAAGTN 1794
Db 841 MAGSGAKSVHYSTYDGGTQ-----GGNYNGD----- 867
Qy 1795 YLTNNPABAI DRINEQGRIRFFHVNDGNQEPVVGQRNGIDSSAGKHSVAIQAKADGEA 1854
Db 868 -----CATGTRSLAVGVGLASAG 887
Qy 1855 AVAIGRQTA--GNOSIAGDQAQATGDSIAIGT-----NVVAKHSG----- 1897
Db 888 ATAVSGAAAGKGSSTAIGRNASADSGVALGDKAGDARGAESYTKYSGLONNVTGT 947
Qy 1898 -AIGDPSIVKADNSVSGNNQFTDAT---QTDVFG-----VGNNI-----TVTESN 1940
Db 948 VSGDASGKETRTVSNAADAKEDAVNLRLDRVAQDANRYVDNKIESLSEGTFFVKVN 1007
Qy 1941 SVALGNSA--ISAGTHA-----GTOAKKSDGTA--GTTTTAGATCTVKGFAGQAVG--- 1989
Db 1008 SL---NNSATPIAGVDATAIGVGATASGASDIAMGNKASASADNAV-----AIGNHS 1057
Qy 1990 -----AVSUGASGAERRIQNVAGEVSATSTDVANGSOLYKATQGIANATNLDRIHQ 2043
Db 1058 VADRANTVSVGSAGSERQVTVNVAAG---TADTDAVNVSQL---NQGLITAKQYTDGVVGS 1111
Qy 2044 NENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQAVAVGLSKLSNGOWVFKIN 2103
Db 1112 LRRDTDGVAATAIATANIPQAYIPGRGMSVGVSSYRQSAIAVGVSSVSGRWVFKFS 1171
Qy 2104 GSADTQGHVGAAGVAGFHF 2122
Db 1172 GSANTRSQVIGAGVGYQM 1190

RESULT 5
AC0976
probable autotransporter sapB [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0976
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0976
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1107 <PAR>
A;Cross-references: GB:ALU513382; PIDN:CAD03303.1; PID:gl6504923; GSPDB:GN00176
C;Genetics:
A;Gene: sapB
Query Match 6.38; Score 670; DB 2; Length 1107;
Best Local Similarity 23.74; Pred. No. 7.3e-19;
Matches 304; Conservative 193; Mismatches 413; Indels 372; Gaps 59;
Qy 993 NSGDDKALI--NAKDIADN---LNTLAGEIRNTKGTAT---DTALQTFQVKKVKGNGDDN 1044
Db 46 NAGNDTGSINQNTTDTATNTTSINNLSNVTTLTDDALLMDAASGTFASR--NGSASK 102
Qy 1045 DADTI--TVGDKAKTNQV-----TULKKGKGLDITQKDGTVTGTGINTQSGKAGNN 1096
Db 103 ITNLAAGTAAADS--TDAVNGSQLFDNTEKVDQNTADITTN-----TNSINQNTTDTATNT 156
Qy 1097 TTLNNGLSIKNTAGNEQIOVGADGVKFAKVNNGVVGAGIDGTTTTRDEIGFAGTNGSL 1156
Db 157 TSINN-----LSNSV-----TTLTDDALLMDAASGTF 183
Qy 1157 DKSPHLSKDGINNAGGKKITNIQSGEIAQNSNDVATGGKIYDLKTELENKISSAKTAQN 1216
Db 184 SASR-----NGSASKITNLAAGTTLAADSTDAVNGSQLFDNTEKVDQNTADITTNTN 234
Qy 1217 SLHEFSVADQGNFTVSNPYSSYDTSKTSVI-----TF-AGENGITTKV-NKGVVR 1267
Db 235 SINQNTT--DIATNTTSINNLSNVTTLTDDALLMDADSGTFSASRNGSASKITNLAAGT 292
Qy 1268 VGIDOTKGLTTPKLTGNNNGKIVIDSON--GQNT--ITGLSNTLANVNDKGSVRTTEQ 1324
Db 293 LAADSTDAV-----NGSOLYETNQKVDQNTSAIDINTSITNLSSD----- 333
Qy 1325 GKIIKDEKTRAASIVDVLVLSAGFNLOGNEAVDFVSTYDVTNFDAGNATTAKTVYDDTSK 1384
Db 334 -----NLSWN-----ETTNSPSAGHSSTTKNTKTNVAAGE 363
Qy 1385 TSKVYVYVNDVDDTTIEVKDKKLVKVTTLTSTGTGANKFALSNAQATGDALVKASDIVAHL 1444
Db 364 LSEBSTDA--VNGSOLFETNEKVDQNTDIAANTTN-----ITQNSTAIENL 408
Qy 1445 NTLSGDITAKAGASQANSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKVAKOKLV 1504
Db 409 NTSVSDINT-----SITGLTD--NALLWDE-DTCAFSANHGSGTSKITNVA----- 451
Qy 1505 AQAQTPDGTLAQMNKVSINKEQVNDANKKQGINEDNAFVKGLEKAASDNKTNAVAATVVG 1564
Db 452 AGALSEDST-----DAVNGSOLYETNQK-----VDQNTS-----AIA 483
Qy 1565 DLNVAQTPLTTFAGDTGTTAKKLGETLITKGGQTDNKLTDNNIGVVGAGTGVTKLAKD 1624
Db 484 DIN---TSITNLGTDALSDWDDERGAFAFSAGTSGTGNKITN-----VAAG-----EIASD 529
Qy 1625 LTNLSNVNAGGTTKIDDKGV---SFDVSSGQAKANT--PVLKANGLDLGGKVISNVGKGT 1679
Db 530 ST--DAIN--GSOLYETNMLISQYNESISQLAGDTSETYITENG--TGKVIKRTNDNGLE 583
Qy 1680 DTD-----ANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRTVTKA 1733
Db 584 QDAYATGATGATAVGYDAVASCAGCALGQNSSSSIEGIA-----LGSSTSNRAITTG 638
Qy 1734 -----GTVLGGKGNNDTEKLTATGQVGVVDKNGNANGDLSNVVVKTKQDGSKKALL 1784
Db 639 IRETATSDDGVVI---GNTTDLRELLGALSUGT--DGESYQITNV-----ADSG----- 684
Qy 1785 ATYNAAGTNTLTNNPABAI DRINEQGRIRFFHVNDGNQEPVVGQRNGIDSSAGKHSVAI 1844
Db 685 -----AQDAVTVRQLQNAIGAIVTTPTPKYHANSTEE-----DSLAVGTDLSLAM 728

Db 1670 IHNYGLTSSNGGSLGVYNYTLISALELLANGANALLATESGLTANRVLNABELFGVGL 1729
QY 1072 DIOTNK-----DGTVT-----FG-----INTQSLKAGNN----- 1096
Db 1730 VDAQNGALTLANGNNRYEGTTVTAGELILGANGAFGOTSLLNIAASGANINGYRQTV 1789
QY 1097 -----TTLNNGLSIKN-----TAGNEQIQVAGQVPAKVNNG 1130
Db 1790 GAVTNSGAVTLNGGVLTSGLLTNGGILDITGGALNLAAGSSSTVAGGLTGAGTLNNGG 1849
QY 1131 -----VVGAGIDGTTIR-DEIGFAGTNGSLDKSKPH-----LSKOGINAGKKITNIQS 1180
Db 1850 DLAVSATNSGLSGQTHIADVASVLTGT-GLTGTSAVEVLGTLLNLCANAA---MTNVL 1905
QY 1181 GETAQNNDVATGGKIYDLKTELENKISSAKTAQNSLHPEFVDAEQNN---FTVSNPYS 1238
Db 1906 GGGVINTAAVT-----LSGNN-----PSGAHQIGTDBELTVGOASN 1943
QY 1239 SYDTSKTSVITPAGE--NGIT--TKVNGVVRVGIDOTKGLTT----- 1278
Db 1944 LGASSATVNLGTLTSHLLNGVSESIANVLGAVAGSTVDIIGGADTALTANNSGFLGOYA 2003
QY 1279 ----PKLTGVNNGKG----IVIDSQNGQNTITGLSNTLANVTNNDKGSVRTTEGKIHKD 1330
Db 2004 LAGNSKLTVASTNNLGASSVALAGAGDTLSLGFNGTFGNSVTGSGVLQVTDDAEVTLT 2063
QY 1331 EDKTRAASI-VDVLASGFNLQ-----GNG-----EAVDFVSTY-----DT 1364
Db 2064 SSGVGNTVKVDIADATILNLDIALFDHVLGTNLNVAKNLATTAFDFGTVGGAFSGI 2123
QY 1365 VNF-----ADGNATTAQVY--DDT-----SKTSKVYDVNVDDTT----- 1398
Db 2124 VNLNTTFALSADNAALATLKLSDSVTTVTGTTDRILHGLDLNGGLTLPDGSPPQSQ 2183
QY 1399 ----IEVKDKLGLVKTTLTSTGTGANK-----FALSNOATGD---ALVKASDIWAHL 1444
Db 2184 ANGVTVDLALNSGTISIITGAGNWEHEPVPVPPNVLLEQDRGDILLQLIDADNVGTNA 2243
QY 1445 NTL-----SGDIOTAKGASQANSAGYVDAGNKVYDSTNKKYQAKNDGT---VDKTK 1496
Db 2244 NDLELMINGTTTISAGQGVQSTVQGGYTVANAT-----HNYGMTSNGSGLYVNYT- 2294
QY 1497 EVAKDKLVAQATPDGTTLAQMNKSVINKEQVNDANKQGINEDNAFVLEKASDNKT 1556
Db 2295 ----LSALELLADGA-----NALLATESGLTANRE 2321
QY 1557 KNAAVT-YGDLNAVAQT-PLTFA-----GDTGTTAKKL-----GET--LTIKGGQ 1597
Db 2322 LNAELSGVGLVDAQNGALTLANGNNRYEGTTVTAGELILGANGAFGQTSLLNIASSA 2381
QY 1598 TDTNKLTONNIGVVAGTDGFTVKLAKDLTNLS-VNAGGTKIDDKGVSFVDSGQAK--- 1653
Db 2382 SANINGYRQTVGAVNTGTVTILNGGELTSTDITLNTGMINVTD-GILNLENGGASSISG 2440
QY 1654 --ANTPVLASGLDLGGKVISNVKG-KQDIDAANVQOLNEVRNLLGLNAGNNDAGNQ 1710
Db 2441 GLTGNCILNIKGGDFTISIDNNGLAGQTNISDGASV-TLNGGTTIIGTGNLS----- 2492
QY 1711 VNIADIKKDPNSGSSNRFTVKAGTVLOGKGNNDTEKLAT-----GGVQVGVDKDN 1762
Db 2493 -SVIDVLGDLNLVADNS-----LANVISGDGTINTATVTLSGNSSFSGAHQIG----- 2540
QY 1763 ANGDLSNVVKTKQDGSKKALLATYNAAGQTNLYL-TNNPAEADRI----- 1807
Db 2541 TNGELT-----VQASNLGASATVNLGTLTSHLLNGVSESIANVLGAVAGSTVDIIGG 2595
QY 1808 -----NEQGRFRFFHVDNGQEPVVQGRNGRIDSSAS-----GKHSVAIGF----- 1846
Db 2596 ADTALTANNSGFLGOVALAGNSKLTVASTNNLGASSVALAGTGTDTLSLGFNGTFGNSV 2655
QY 1847 -----QAKADGEAA-----VAICRQTAQNGQSIAGDNAAQ 1876

Db 2656 TGSGLVQVTDAAEVLTSSNGVNAVITDIDATNLDDIALFNHALTGGLNVAKN-- 2713
QY 1877 ATGDQSTAGTGNWAGKHSGLGDPSTVKADNS-YSGVNNNOFTDATQT-----DVFG 1929
Db 2714 --DASTAPDFGATVGAFGTG-----TVNLNNSTPDLSCGNTTTLAQAATLKLSSGNLTS 2764
QY 1930 VGNNI-----TVTESNSVALGSNSAISAG--THAGTQAKSDGTAG----- 1968
Db 2765 VGVQVQNIQTLAMNGGTLFDNIVDNRAGIITSDGTTAANSINTTGGGEVRVNLPSNLA 2824
QY 1969 -----TTTTAGATGT-----VKGFAQGTAVGAVSVCASG 1997
Db 2825 LDGLSVNMLDEBEIIVTLATGAATGTGHELTLDENGDDPSAVTYQGVHNGSTSAATG 2884
QY 1998 AERR-----IQNVAAAGEVSATSTDAV-----NGSQLYK----- 2025
Db 2885 SFNYGMTTGEDYDGLYVNYGLTALELLSTGSEALVLTAILANNGTQSDNLSAQITGSGDL 2944
QY 2026 -----ATQGIANATNEL-----DHRIHQENKANAGISSAMAMASMPQAYIPGRS 2070
Db 2945 AFASANDGSTASLSNSTSYTGTTWSSGNLRLDADSA-LGQTSLLAMSTATHVDINGTQ 3003
QY 2071 MVTGGIATHNG-----QGAVALGSLKSDNGQWVFKINGSADTQCHV 2112
Db 3004 QVVELATEGGSTLDLNDGKLTVTGCGQIDGALTGGGELVLSGS-----LNVSYDNAGFT 3059
QY 2113 GA---AVGAGFH 2121
Db 3060 GSTDIANGAVAH 3071
RESULT 7
B71704
cell surface antigen (sca3) RP451 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: B71704
R:Andersson, S.G.B.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:95039499; PMID:9823893
A:Accession: B71704
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2340 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:G3868717; PIDN:CAA14908.1; PID:G38610
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: sca3; RP451
Query Match 6.1%; Score 651.5; DB 2; Length 2340;
Best Local Similarity 21.2%; Pred. No.1e-17;
Matches 519; Conservative 302; Mismatches 879; Indels 743; Gaps 121;
QY 10 NKATGTFMAVAE-----YAKSHSTG-----CGSCA-----TGQVGSVRTLSFARIAALAVL 55
Db 4 SKILRKFLATASLCGTLFTNSATGTIIPNNGSVSLNTDAGLVGV-----FNNGDIQI- 58
QY 56 VIGATLNGSAYAGIGISEADGGKGANARGDKSIAIGDIAQALGSQSIAT----- 105
Db 59 -----VNGGREIKISADKANAIIGINTLKE-----LPDFGGEVVSQVYISGPLNAGEDLN 109
QY 106 ---GDNKTVHNS-----NNANIGAK-----ASGNSI 130
Db 110 TNFGPLKFIENNVTIITGVGKTFPSNIDFAKNATLQINKDLNITTTIDNTVAGNNGSI 169
QY 131 AIGGDVLASGH-----ASTAIGSDDLYLKKEVTQQLISELLPIIRGQKALNDIYQLADT 183
Db 170 TFGSGIISNHIQYTNLSLLGVNGGEAKIYAPEANNIT-----I 209
QY 184 NLOKYRRTHAQCHASTAVGAMSAYAKHFSNAGFTGRATAGTYSLAVGLTATAKAASSIAV 243
Db 210 NAKNINLTNNSILTLCDGNITTLKGNINNT--TEIDQGGILNLAAYDL-----GSSSIIT 262

QY 244 GSNQAIGPAATAVGSGTQVNLNRGIALGFGSOVLQKQNDVNAANVRAYAPDDNQPIDNR 303
 Db 263 GD-----IG-----NIGSLDITVILGSGAT-FNSTILKATNINLKHTSTLNLDDN----- 307
 QY 304 YKATFKNGATDFVSI GNSGNDSSIRKIIIN--VGAGSADTDVAVVQALKEAVELANROIT 361
 Db 308 -----IIVIGNIKGNNN--XDLNFKVHGHTNLDNEMIIIPAPQK-----THGTLN 349
 QY 362 FKDDSNV-----RVEKGLGKLTITGGAQ-----TSA 399
 Db 350 FKGNATLNGNINNLNLIKFSGGHGKTLNQGNTKVDNLVPADSVLDSGTTISVNGLLDTC 409
 QY 390 LTDHNLGVWONGGLVQLAETL-----TSLKMTTENTLANTANEKVTVGKTRLTLDKIG 442
 Db 410 VTFNNNV--NGTLLINAKNTISAKLNAKAKIQINANLWNH-----PSAG 456
 QY 443 FTMNMGIDESKPYLDRDTGIHAGGOKITKLTAGVVDDDAATYGQLKKVNTQAESALQTF 502
 Db 457 DISDIRIADNTIYTIDAKN-----GNVLLNNAKIIIFEGADS--MLALIN-TGVTDARTF 509
 QY 503 TV-KKVDKNGNDAND-----SKITVGNKNKPDGT-QVNTLKLK----- 539
 Db 510 TIYNLNQSGNDEYIGVIEAKKVIITIANQSGPYTIGQDNTHLRELIVEGAGDIIIDD 569
 QY 540 -----GENGVDDVTETNGTGTVTFGLNQNNGLTVGN 568
 Db 570 TIFTKLLSINSTQITFNRTLDLGAGNIAFGHGLVLVNGVTGSIITSENNOGILTIIS 629
 QY 569 STLNDGLSVKNTNSNKQIQVGADGIT-----FTDISNSKPGA-----GI----- 608
 Db 630 G--NITGVIGTNELGLKLVNIGADPVCNSANVPASVALTNPSSVLIILADGVTLTGEVTH 687
 QY 609 ENTRITRDIIGPANNITGSLDANKPRLTPPGINAGGKELTNVQSAINPATNGGQLODFMR 668
 Db 688 NNTKGVLISLGTG--SNITGQIGTNSAALKINIGAGA---SNIDSNIYAG----- 732
 QY 669 LSPANTEKSSAATIKDLYNLSOVPLTFAGDTGPNVTKKLGEILKVKGGKTTA--DDLTKN 727
 Db 733 -STVLTDQISELTLNNDVVVNSNIIITTAGNNSKLIPTGNGGIGTGNIGANGAALQEVFN 791
 QY 728 ---NIGWADSTDSNLTVKLAK-----TLSDLDVANTKLTITASDKVT--VDSGNNTAKL 776
 Db 792 GTTINIGGTANSQNTVAHSAANVVITGLTGTALKYKXDTGIIAHGGLVGDIDFNKNKAGP 851
 QY 777 QNGD-----LTF-----SKQNTGATPATNSKTIG----- 800
 Db 852 ILGDGAMIDGSLCNGVAGTLDFIGDGNVTQNIAGDANANSISTINIQGDNTKNVTIAND 911
 QY 801 --VDGLKFTDNNGIALDGTYYITKDKVGFAPQDGLDKSPYLDKDKLVGVEVEITNGI 858
 Db 912 IFVDNTHFTNGGILQGGN--LTHNIDFGANGGTLFN-----GNNTYLNLA 958
 QY 859 NAGGKAITGLSNLTLD--ATNATTG-----HVTQLGI-----VSDTKTRAASIGDVL 904
 Db 959 IVNGQ--NGLNAFTNLKASDDTIGTVKIINIIGQIGTPQNTIIVANNKLTIVSSVSSI 1016
 QY 905 NAGFNLKNGDAKD-----FVSTYDVTDFPING-NATTAKVTYDGKASKVAVDVNVDGTTIH 959
 Db 1017 NF-----GDANSQILSAPVDQTIKFNNLNETGGI-----ITLDSN--GNLNT 1059
 QY 960 LTCADGNKNOIGVKTTLTKTDAGDKAI-----NFSVNSG--DDKALINAKDI 1006
 Db 1060 ISGNNGIK--LGSKNELSLNIGKVTVNDLIDIQNIHQINNGALFDDQSLTSKAKIK 1117
 QY 1007 ADNLNTLAGIRNTKGTADTALQTFQVKVKENGDDNDNA-----DTITVGKDAKTNQV 1060
 Db 1118 NINIGTVAGG-----ATYTLDAINDFNOLNTSGWFKHQDSILLEKNSNTND 1165
 QY 1061 NTULKKGKNGLDQTNKDGVTFTGINTQSGLKAGNNTLNNGLSINKNTAG--NEQIQVG 1118
 Db 1166 HTITL--TSALDPGNNGQGIKLTIDT-----NKLITIDNNG-NVAYTILGTANHMLK-- 1213

QY 1119 ADGVKFAKNVGV-----VGAGIDGTTRITRD-EIGFAGTNGSLDKSKPHLSKDGINA--- 1170
 Db 1214 --QLTFASIDNGAIALKVGINVENTVNLIKDIELNEVNAVNLFNKNTYTTATGNGHVD 1271
 QY 1171 --GGKKITNIQ-----SGEIAQNSDAVTG-----GKIYDLKTELENKISSAKTAQAN 1216
 Db 1272 FQGNAGVINLNDIDIEDGSVTSGN--VNGTLNFGSGKVTGLNNIIVMLQAGAGDVLS 1329
 QY 1217 SLHFSVADBOG---NNFT-VSNPYSSYDTSKT-----SP 1247
 Db 1330 ASGNSITBIQNGNNNLTPAANSHTTIDINKTGGQDLNLVINGSGSVSSIGANAAGVD 1389
 QY 1248 VITFAG-----ENGITTKVKNKGVVRVID--QTKGLTTPKL----- 1281
 Db 1390 IINAGSVNFSNTLKSNGNIVISDGAITQVNNVNTATDISKQANNGLTKLNNHTPINITS 1449
 QY 1282 TVGNNGKGVIDSQNGQNTIT-----GLSNTLANVTNDKGSVR 1320
 Db 1450 TLGNNAIG--TIEVANNNDVTITGTLOAQNIHFSNATQAATLTILGAASQVTNITTAGNNIH 1508
 QY 1321 TTEQGKLIKDEDKTRAASIVDVLS--AGFNLOQNG-----EAVDFVSTYDVTNPAAGNA-- 1372
 Db 1509 TLE-----VTDFTDNDGIIIGDANNRLKSIELTNGTVTINSPHYSSITANNNAQGNVXL 1564
 QY 1373 -TTAKVTYDOSTKTSKWVDVNVDDTTI--EVKDKKLGKVTTLTSTGTGANKFALSNOA 1429
 Db 1565 NIEGGITYDLGSKIKSLANVOISEDITIRGDVYSKVLNIDAGKTINFRDGNMNNPKND 1624
 QY 1430 TGDALV-----KASDIVAHLNLTSGD--TOTAKGASQAN--SSAGYVDAD--GNKVIVDSTD 1480
 Db 1625 IPDALIDLVLPRSLSLFNFTDIKADNLNFADDTATANKDAVWIDAHIDNGGILKFND 1684
 QY 1481 NKY--YQAKNDGTVDTKEVAKDKLVAQAQTPDGTGLAQMNKS-----VINKEQ 1527
 Db 1685 NAWLTOEIKNANTII-----ETASDKF-----LLQKNIKATLIADNANILVLDNVE 1731
 QY 1528 VNDANKKQGINEDNAFVKLEKASDNKTNAAVTVGDNLNAVAQTPLTFAAGDTGTTAKKL 1587
 Db 1732 VNTNLNVRDIVLDA--NYELKYTGNTVTHNGLLTI-----ITY--PDALQKG 1775
 QY 1588 GETLTIKGGQTDTKLTDNNIGVAGTGTGTVKLAKDLTNLNS-----VNAGGTK 1637
 Db 1776 GHILVSGSGNVMSDL--DNLI-----IKIKAHSDITNITSDTKHQIVKLETGAIYTP 1826
 QY 1638 IDDKGVSVFVDSQOAKANTPVLISANGL-----DLGGKVISNVG-----KGTKDTDAANVQ 1688
 Db 1827 VPQTKVIIIDASEBQNKVFWADANGLVLLTDTGGR--DDTGRDDTGRGNTD----- 1878
 QY 1689 LNEVRNLLGLGN-----AGNDNADGNOVNIADIKK--DPN-----SGSSS 1726
 Db 1879 -NGCRDNCVGNLSNNSNEAGSSSDKN--YGITDVVPFDPSPILDYTKNNYVAGSIAN 1936
 QY 1727 ---NRTVIKAGTVLGGKGNNDTEKLATGVGVQVQKDG---NANG-----DLSNVWVK- 1773
 Db 1937 QLINHVKDFGNTTDAGKLLNDLGFMSPNRVTETLDRLSNRINVLNGLNEGVGLNGIEVEN 1996
 QY 1774 -----TQKD-GSKKALLA---TYNAGOTNYLTNNPAAEADRINEGIRFFHV 1817
 Db 1997 FLTDIAINMDNFTAKSIGRLEBELSDANTVGNLTKNTLNKK-----INLKL 2045
 QY 1818 NDNQNPVQVRNGIDSSAGSKHSVAIFQAKADGAAVAIGRQTOAGNQSIAIGNAQA 1877
 Db 2046 NTNNOAIAGAEB--DNIVTIGWMSFYCKIKQNSKNS--ASGVQSNWGGGIIIGFDYNI-- 2100
 QY 1878 TGDQSIAITGNVAVG---KHSGAIGDPTVKADNSYSVG-----NNNQFTDA----- 1922
 Db 2101 --DNSIVIGAAVTMADSKVKKNDKNGDRTKAKSNISYISYGLYNLNTNFFVEAIGVYGR 2158
 QY 1923 -----TQTDVFGNNITVTESNVALGSNAISAGT 1954
 Db 2159 NKIKNYEKRIITITDQIAGIKFINTFYSELLGGYNYLISHRT 2201

```
RESULT 8
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R:Hayashi, T.; Makino, K.; Kurokawa, M.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033965.1; PID:gl3360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: ECs0542

Query Match 6.0%; Score 645.5; DB 2; Length 5291;
Best Local Similarity 21.9%; Pred. No. 5.3e-17;
Matches 582; Conservative 293; Mismatches 932; Indels 855; Gaps 131;

Qy 35 TQGVSVRTLSFARIALAVLIVGATLN----GSAYAGIGISEADGG-KGGANARGDKSI 89
Db 1136 SGPTLTINTVSGDDIINAAETVVAQTISGQVGTGAVAGTIVITIGGNQYNATVQSDLSW 1195

Qy 90 AI----GDIQAALGSQSTAIGDNKIVHNNANIGAKASNESIAIGDVLASGHASIAI 146
Db 1196 SVSVPANVLQALNGELTISAS--LTNSANN-----TGTATHDIVIDANLPLGLRVDV 1246

Qy 147 GSDDLKKEVTVQOISELLPIRGOKALNDIYQLADTWLQVRRTHAQGHASTAV---- 201
Db 1247 AGDDVINSIEHTQAL-----VITGSS-----GLAAGALTIVINSVT 1284

Qy 202 -GAMSYAKGHFSNAPGFRATAE---GTYSVLAVGITATAKAASSIAVGSNAQAIGFAATV 257
Db 1285 YGATVLADGWSVGVPPVADVNTWEPAGTVNIAVSGTNTAGTTTSI---SHPVTVDLAAVAI 1341

Qy 258 -----CGSTQVNLNRGIALGFGSQ-----VLQKN-----DVNA 286
Db 1342 TINTLSTDDVINAAEKSGDQLSGTTSGEVAGQITVIFGKGSYTTTVAADNTWGLTIPA 1401

Qy 287 ANVRAPYDDNQPIDNRKATPKNGATDV-FSIGNSGNDSIRKKINVGAGS----- 338
Db 1402 VDV-----ATLPGAANVQASVSNVAGNSQTATHASVDATPSVTINT 1445

Qy 339 -ADTAVNVAQLKEAVRLANRQ-----ITPKGDDSNRNRVKGGLKTLITITGG---- 384
Db 1446 IATDDILNAAEAGSALTISGTSTAEAGQTVTVTLNGVNYSGNVQADGWSVSVPTGDLAS 1505

Qy 385 -----AQTSAITDHNIGVQNGDGLKYQLAETLSLKKVYTNEL----- 423
Db 1506 LTASSTYTNASVSKARNSASATHN-----LTVDLAAPVVTINTVAGDDIINATEH 1556

Qy 424 -----TANE-KVTVGKRLITDKIGFTNDMNGIDSKPKYLDKDTGIGHAGQ 468
Db 1557 GQAQIISSGATGATTGNVTSTVITGTYTT-----VLDANG-----NWSIGVPAS 1601

Qy 469 KTKLTLTAG-----VVDODAATYQLKKVQNTAES-----ALQPTTVKKVDRKNGNDANDSK 518
Db 1602 VISALAQDVTITATVTDSDAGNSGTASHTVTVVALGAPVLAINTIAVDDIINAAEKADLA 1661

Qy 519 IITVGNKPKDGTQVNTLKLKGENGVDVTTETNG--TVTFGLNQNGI-----TVGNSTL 571
Db 1662 I--TGTSNQPAQTQI--VTLINGQN--YTTTADAGNWSVTPASRVLSALGEATYTTVTAAT 1717

Qy 572 NNDGLSVKNTSNKQIQVGADGITF-----TD-----ISNSKPGAGIENTTRI 614
Db 1718 DADGNS-GSASHNVQVNTALPGVTINVVATDDIINAAEAGVEQITISQVGTGAAGDVTV 1776

Qy 615 TRDGIGFANNVTSGLDANKPLTPTGINA-----GKELTNVQSAINPATNG----- 661
```

Db 2793 -----VGVPADVTALGSGAQTTITASVSDRAGNSDDASRTVTIVLSAPV 2836
Qy 1442 AHLNLTSGD---IQPAKAGSAQNSAGYVDADGNKVYDSTDNKYQAQKNDGTVDKTEV 1498
Db 2837 ISINTIAGDDVINATEKSGDLALS--GTSQDPAGTAITVTLNGQNTSATTTDASGNMSVTV 2894
Qy 1499 AKDKLVAQAQTPDGTGLAQMNKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDNKTKN 1558
Db 2895 P-----ASAVSALGEATYSVTASVTN-----AQGNSSTASHNVQVN 2930
Qy 1559 AAVTVGDLNAVAQPLTFAGDGTGTTAKKLGELTLIKGGQTDNKLNDNNIGVVGAGTDFGT 1618
Db 2931 TALPGITINPVATDDIINASEAGSAQTISQVGTAAAGSTVTVEL-----GCKTVT 2981
Qy 1619 VKLAKDLTNLNSVAGGTKIDDKGVSFVDSGQAKANTPVLNGLDLGCKVLSNVKGT 1678
Db 2982 ATVQADLSNWSVPAADWQ-----ALNGELTIVNASVTNA-----VGNTGSGT 3024
Qy 1679 KD--TDAANVQQLNEVRNLLGL--GNAGNDNADGNQVNIAD--IKKDPNSGSSNRTVIK 1732
Db 3025 RLITIDA-----SLPGLRVDTVAGDD-----VNMIEHAQAQVITGSSG--FA 3066
Qy 1733 AGTVLGGKGNNDT---EKLATGQVQVGVQDKGNANGDLNVMVVKTKQKGSKKALLATYNA 1789
Db 3067 AGTALTIVINNQTAAATVLANGWSVGPAT-----DVSN-W---PAGTLNITVSGANS 3116
Qy 1790 AGQTNVLTN-----NPAAIDRIN--EQGIRF-----FHVNDGQEPVVOG 1828
Db 3117 AGTQTSITREPLTVDLPAVAISMNSITSDAINAAEKGAALTLSGSTSGVEAGQTVTVTFG 3176
Qy 1829 RNGIDS--SASGRHSV--AIGFQAQADGEAAVAIGROTQAGNQSTAIADNAQATQDSI 1883
Db 3177 GKYTTTVAANGSWSTVTPAADLAALRDGDSAQV--RVTVN-----GNSATATHEYSV 3229
Qy 1884 -----AIGTGNVAKHSGA---ICDPSTVKADNSYSV---GNNNQTFDATQTDV 1927
Db 3230 DSAAPTVTINTIASDNIINASEAAAAGTVTSGTSTAQTGOTLTVTLNGTVQTT--VQTD- 3286
Qy 1928 FGVGN-NIIVTESNVALGSNSAISAGT--HAGTQAKSDG-TAGTT-----TTAG- 1974
Db 3287 ---GWSLTPASDLTALANNGYTLTATVSDLAGNIGSAGKGVTVDTTAPVIFSNFVAGD 3343
Qy 1975 -----ATGTVKG-----FAGQ-----TAVGAVSVGASGAERIQNV 2005
Db 3344 DVINNVEHIOAQIISATGAVAGDRVLVTIAGQQVSTSDAGNSVGVPASV--ISGL 3401
Qy 2006 AAGEV--SATSTDVANGS-----QLYKATQGIANATNELDRIHQENKKNAGISSAM 2056
Db 3402 ADGTVTISATITDSAGNSSTQTHNVQVNTAAVSLSVSTISGDNLI---NAABAG--SAL 3455
Qy 2057 AMASMPQAVIPGHSMTVGGIAT--HNGQGAVALGSLKSDNGQWVFKI-----2102
Db 3456 TLSG-----TGTFNFGTGVTVTVLLNGKGYSAT---IQSNGSWSNVPAADVAAALSDGT 3505
Qy 2103 -----NGSADTQGH 2111
Db 3506 STVTSASAQDSAGNSSTQTH 3527

RESULT 9
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
R;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; PMID:12074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-5188 <STO>
A;Cross-references: GB:AE005174; NID:gi2513368; PIDN:AAQS4838.1; GSPDB:GNO0145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0615

Query Match 5.9%; Score 634; DB 2; Length 5188;
Best Local Similarity 22.0%; Pred.No.1.5e-16;
Matches 576; Conservative 290; Mismatches 928; Indels 830; Gaps 130;

Qy 35 TGOVGSVRTLSFARIAALAVLVIGATLN---GSAYAGIGISEADGG-KGGANARGDKSI 89
Db 1136 SGPTLTINTVSGDDIINAAEIVVAQTISQVGTGTAAGTAVTIVTIGNQVNAIVQSDLSW 1195
Qy 90 AI---GDIAQALGSGSIAIGDNKIVHNSNNNINIGAKSGNESIAIGDVLASGHASIAI 146
Db 1196 SVSVPANVLQALNGELTISAS--LTNSANN-----TGTATHDITVIDANFLGLRVDTV 1246
Qy 147 GSDDLVLKKEETVQOISELLPIIRGQKALNDIYQLADTNLQKYRRTHAQGHASTAV--- 201
Db 1247 AGDDVINSIEHTQAL-----VITGSS-----GLAAGAALTIVVINSVT 1284
Qy 202 -GMSYAKGHFSNAFCTRATAE---GTYSIAGVLGTATAKAASSIAVGSNAQAIIFAATAV 257
Db 1285 YGATVLADGWSVGVFPVADVTPNWPAGTVNIAVSGTNTAGTTTGI---SHPVTVDLAVALI 1341
Qy 258 -----GGSTQVNLNKGIALGFGSQ---VLQKDN---DVNA 286
Db 1342 TINTLTDDVINAAEKGSDQLSGTSGVEAGQTTIVFGKSYTTTVAADNTWGLTIPA 1401
Qy 287 ANRVAPADPNQPIDNRYKATFKNGATDV-FSIGNSNGNDSIRRKIINVGAGS----- 338
Db 1402 VDV-----ATLPDGAANVQASVSNVAGNSQATTHAYSVDATAPSTINT 1445
Qy 339 -ADTDVAVNAQLKEAVRLANRQ-----ITFKGDDSNNRVERKGLKTLITGG----- 384
Db 1446 IATDDIILNAAEAGSALTISGTSTAEAGQTVTVTLNGVNYSGNVQADGWSVSVPTGDLAS 1505
Qy 385 -----AOTSALTDHNIQVQNGDGLKVQLAETLTSUKMYTTEL-----423
Db 1506 LTASSYTVNASVSDKARNASATHN-----LTVDLAAPVTVTINTVAGDDIINATEH 1556
Qy 424 -----TANE-KVTVGKTRLTDDKIGFTNDMNGIDESKPYLDKDTGIIHAGGQ 468
Db 1557 GQAQIIISGATGATGNTVSVITGTTTYYT-----VLDANG-----NWSIGVPAS 1601
Qy 469 KITKLTAG-----VVDDDAATYQQLKKNQTAES---ALQFTYVKYVDKNGNDANDSK 518
Db 1602 VISALAQGDVTTITATVTDAGNSGTASHTVTVALGAPVLAINITIAVDDIINAAEKGADLA 1661
Qy 519 IITVGNKPKDGTQVNTLKGENGVDVVTETNG--TVTFGLNQNNGL-----TVGNSTL 571
Db 1662 I--TGTSNQPAQTQI--TVTLNGQN-YTTTADAGSNMSVTPASRSVAGALGEATVTVTAAAT 1717
Qy 572 NNDGLSVKNTNSKQIOVGADGITE---TD-----ISNSKPGAGIENTTRI 614
Db 1718 DADGNS-GSAHNVQVNTALPGVTINNVAVDIDIINAAEAGVEQITISGQVTGAAGDVTIV 1776
Qy 615 TRDGIQFANNITGSLDANKPRLTPTTGINA---GGKELTNVQSAINPATNGG-----661
Db 1777 T---LGGATYVATVQANLSWSDVPASALQELNGELTISASVTSVNGTNGTREITID 1833
Qy 662 -----QLDFMRLSTANTEKSGSAATI---KDLNLSQVPLTFAGDT-----700
Db 1834 ANLPGLRVDTVAGDDVWNIIIEHQALVITGSSSGLAAGSNVTLITINGQTVYVAVLADGTW 1893
Qy 701 ---GNVTVKK---LGEILLKVGKGTAD-----DLTKNNI---GVVADSDTNSL 740
Db 1894 SVGVPAADVSAWPAAGSVTTIAASGNSAGNPVSVTHPVTVVDLSAVAVSINAITADDVINAA 1953
Qy 741 TVKLAKTISDLDAVNTKTLTASDKVTVDSGNNT-----AKLQNGD 780
Db 1954 EKGAAALTLSG-----STSGVEAGQTVTVTFGGKTYSATVAANGSWTSVPAAQWALRDGD 2009

A;Reference number: A89758; MUID:213111952; PMID:11418146
A;Accession: B89921
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-6713 <R>
A;Cross-references: GB:BA000018; PID:g13701232; PIDN:BA842527.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: ebha

Query Match 5.8%; Score 622; DB 2; Length 6713;
Best Local Similarity 20.5%; Pred. No. 6e-16;
Matches 559; Conservative 317; Mismatches 970; Indels 876; Gaps 124;
QY 6 KVPFNKATGFMVAEYAKSHSTGGSCATQGVGSVRTLSFARIAALAVLIGATLNGSA 65
DB 39 ETILNKOTGP--NTAKTAVEQALNNVNSAKHALNGTQNLNAKQAATAINGASDLNQK 96
QY 66 YAGIGISEADGGKGGANARG-----DKSIAIGDIAQALGSQSAIGDNKIVHNSNNAN 119
DB 97 KDALK-AQANGAQRVSNANDVQRNATELNTAMGQLQHAIAKTKNTLASSKYVNADSTKQ 155
QY 120 -IGKASGNESIAIGDVLASGHASIAIGSDDLKLYKETVQOISELPIIRGQALNDIY 178
DB 156 AYTIVTNAEHIISGTPVTTPSEVTAAN-----QVNSAKQELUGDBRLR--- 202
QY 179 QLATNLQKYRRTHAQGHASTAVGAMSAYAKGHFSNAFGTRATAGTYSLAVGLTATAKAA 238
DB 203 -----VAKQNTAIDALTQ-----LNTPOKAK 225
QY 239 SSIAGVNAQAIGAFAATAVGSGSTOVNLRGIALGFGSOVLQKDNVDVNAANVRAY-----A 293
DB 226 LKEQVGQANRLEDVQSVQTNQSLNNAMKGL-----RDSIANETTVKASQNTYDA 275
QY 294 PDDNOPIDNRKATFKNGATDVFSIGNSGNDSIRKLIIVGASADTDANVAQ--LKE 351
DB 276 SPNNQ-----STYNSAVNSAKGIINQTNPT-----MDTSATQATQVNNAKNGLNG 323
QY 352 AVRLANRQITPK-----GDSNNRVEKGLGLTLITGGAQT 387
DB 324 AENLRNAQNTAKQNLTLNLSHLTNNQKSAISSQIDRAGHVSEVTAAKNAATELNAQMNL 383
QY 388 SALTDHNIQVQNG-----DGLK-----VOLAEFL-----TSLKMWTTENITAN 426
DB 384 QAIHQDQ--TVKQGVNFTDADKAKRDATYNAVSRAETILNKTQANTSKQDVE-----AAI 437
QY 427 EKVTVGKRLTIDKIGFTNDMNGIDSKPYLDKDTGHHAGGOKITKLTAGVDDDAATYG 486
DB 438 QNVTSAKNALGDO-----NVTNAKNAKALNNUTSIN-NAQK-RDLTTKI--DQATTVA 489
QY 487 QLKXV-----QTAESALQTFVKKVD-----KNGNDANDSK-----IITVGKN--NKP 528
DB 490 GVEAVSNTGTQNLNTAMANLQNGINDKANTLASENYHDADSKKTAYTQAVTNAENILNKN 549
QY 529 DGTQVNTLKLKGVNDVTTETNGTVPGLNQNGLTVGNSTLAN--DGLSVKNTNSKQI 587
DB 550 SGNLND--KAAVENALSOVNTAKGALNGNHNLEQAKSNANTTINGLQHLITTAQDKLQK 607
QY 588 QVGADGTTFTDISNKPQA--GIENTTTIT--RDGIFPANTGSDANKPRLTPTGINAGG 644
DB 608 VQOQNAGVDVTVKSANTLNGAMGLTNSIQDWTATNGQNYLDATF----- 655
QY 645 KELTNVQSAINP-----ATNGGOLDPMNRLSTANTEKSGSAATIKDLYNSQV--PLTF 696
DB 656 RNKNTYNNAVDSANGVINATSNPNMD--ANALNQIATQVTSKNAIDGTHNLTOAQQTATN 714
QY 697 AGDTGPNVTXKLGLILK-----VKGGKTTADDLTKNNGIVGVADSTDNSLTVKLA 745
DB 715 AIDGATNLNKAQKALKAQVTSARVANVTISIQQTANEL--NTAMGQLQHGIDDENATKQT 773
QY 746 KTLSDLD-----AVNTKTLTASDKVTVDSG-----NNTAKLQNGDLTF-- 783
DB 774 QKYRDAQSKKTAYDQAVAAKAILNKTQSGNSDKAAVDRALQVVTSTKQALNGDAKLAE 833

QY 784 ----SKQNTGA----TPATNSKITG-----VDGLKFTDNNGIALDCTTYITKDKVGFA 828
DB 834 AKAAARQNLGLTINHITNAQRTALEGOINQATTVGVNTVKTNTANTLDGAMNSLOGA-- 891
QY 829 KQGSLSKSPYLDKDKLKVGEVEITNNGINAGKATIGLSNTLTDTATNATTHVHTQGI 888
DB 892 DKDAITL-RNQNYLDADESK-----RNAVYTAQVTAABEILNKQGTGNTSKADVDNA 940
QY 889 VDSITDKTRAASIGDVLNAGFNKNGDAKDFVSYDVTVDINGNATTAKVYDYGKASKVA 948
DB 941 LNAVTRAKAA-----LNGAENLRN-----AKTSATNTINGLPNLTQOLKMLKHQVE 987
QY 949 YDVNVDTTIIHLTGADGNKQ-----IGVKTITL-----TKT-----DAKGDKAIFNS 991
DB 988 QAQNV-----VGVNGVKDKGNTLNTAMGALRTSIQNDNTTKTSQNYLSDASDSKNKNYN 1040
QY 992 VNSGDDKALINAKD-----IADNLMTLAGEIRNTKGTADTALQTFQVKKVKGEGDDDDAD 1047
DB 1041 TAVNANGVINATNPNNDANAINDMANOVNTTAAALNGAQLAQ----- 1085
QY 1048 TITVGKDAKTQVNT-----LKLKGKGLDIOQKDTGTVTFGTGINTOSGLKAGNNTLNN 1101
DB 1086 -----AKTNATNTINNAOPLNQKQKDALKTQVNNAQRVSDANNVQH-----TATELNG 1133
QY 1102 NGLSIK-----NTAGNEIQIQQGADGVKFAKVNNGVVVGAG--IDGTTTIT----- 1143
DB 1134 AMTALKAALADKERTKASGN--YVNADQEKQAYDSKVITNAENIINGTPNATLTVDVN 1190
QY 1144 --RDEIGFAGT--NG--SLDKSPHLSK--DGINAGGKKTINOSGEIAQNSNDVATGGK 1195
DB 1191 SAASQVNAAKTALNGDNNLRVAKEHANNITDGL-----AQLNNVQAKLKEQVQSATTLDG 1246
QY 1196 IYDLKTELENKISSTAKTQNSLHFSVADE--QGNFTVSNP--YSSVDTSKTS-- 1246
DB 1247 VQTVKNS-----SOTLNTAMKGLRD--SIANEATIKAGONYIDASFNNRNEYDSAVTRA 1300
QY 1247 -----DVITFAGENGITTKVNGKGVVRVGIDQTKGLTTPK-----LTVGNNGK 1289
DB 1301 IINQTSNPTMEPTITQA--TSQVTTKEHALNGAQLAQAK--TTAKNNLNIITSINNAQK 1357
QY 1290 GIVDSQNGQNTITGLSNTLANVTNDKGSVTTTEQGIKIDEDKTR----- 1335
DB 1358 DALTRNDGATTVAGVNGQETAKATELNNAMHSLQNG--INDETQTKQKYLDAPESKKS 1415
QY 1336 -----AASIVDVLSAGEN-----LQG-----NQE-----AVDFVSTY 1362
DB 1416 AYDQAVNAKAILTKASQGVNDKAAVEQALQNVNSTKTALNGDAKLEAKAAKQTLGTL 1475
QY 1363 DTVNFADGNATTAKVTYD-----DTSKTSKVVYD-----VNVDDTTI-----EV 1401
DB 1476 THINNAQRNALDNEITQATNVEGVNTVKAKAQQLDGAMGQLETSSIRKDKITTLQSQNYDA 1535
QY 1402 KDKKLGVKTTTLTSTGTGANKFALSQATGDALVKASDIYVAHLNLTSGDITQAKAGAOAN 1461
DB 1536 DDAKRTAYSQAVNAATAILNKTAGGTPKAD--VERAMQAVTOANTALNGLQNLERAKQAA 1594
QY 1462 -----SSAGVYVDAG-----NKVIYSTQNKY 1483
DB 1595 NTAITNASDLNTKQKEALKQVTSAGRVSAANGVEHTATELNTAMTALKRAIADKATK- 1653
QY 1484 YQAKNDGTVDKTKEVAKDKLVAQAA-----TPDGTTLAQMNVKS-----VIN-KEQVN--DAN 1532
DB 1654 -ASGNYVNAQDANKRQAYDEKVTAAEHIVSGTPTPLTSPDVTAATQVNTAKTQLNGNHN 1712
QY 1533 KKQGINEDNAFVKGLEKAAASDNKTK-----NAAVTVGDLNVAQOTPLTFAGDTGTTAKKL 1587
DB 1713 LEVAKQNTAIDGLTSLNGPQKAKLKEQVGQATTLNPNVQTRDNAQTL-----NTAMKGL 1768
QY 1588 GETL-----TIKGGQDGT-----NKLTDDNNIGVVA-----GTDGFTV 1619
DB 1769 RDSIANEATIKAGONYTDAQNKONDYNNNAVTAAKAIIQOTTSPSMLAQEINQAKQVTA 1828

```
QY 1620 KL-----AKDLTNLSVNA-----GQTKIDD-----1640
Db 1829 KOQALNGQENLTAQNAKHLNGLSLDITNAQKDAKROGATHVNEVTOQNADALN 1888
QY 1641 -----KGVSVFDSGGQAKANTPVLISANGLDLGGKVS-----NVGKGT 1678
Db 1889 TAWTNLKNIGIQDNTIKQGVNFTDAD-EAKRNA---YTNNAVTAQAEQILNAQKQNPATKDG 1944
QY 1679 KDTDAANYQOL-----NEVRNLLGLGNAG-----1702
Db 1945 VETALQNVQRAKKNELNGNQNANAKTAKNALNLTSSINNAQKAAKLSQIEGATTVAGVN 2004
QY 1703 -----NDNAD-----GNQVNTADIKKDPNSGS 1724
Db 2005 QVSTWASELNTAMSNLQGINDEAATAKAAKYTEADRDQTAYNDAVTAATKILLDKTAGS 2064
QY 1725 SNRTVIK-----AGTVLGGKGN-----NTEKLAATGG 1752
Db 2065 NDNKVAVEQALQRVNTAKTALNGDARLNEAKNTAKQOLATMSHLTNAQKANLTEQIERGT 2124
QY 1753 VQGV-----DKOG-----NANGDLNWNVVKQKDGSKKA 1782
Db 2125 TVAGVQGIQANAGTLNQAMNLRQSTASKDATKSSDYQDANADLQNAV-----ND-----2175
QY 1783 LLATYNAAGQTNLYLTNNPAEAIDRINEQIRFPHVNDGQEPVVOGRNGIDSSASGKHVS 1842
Db 2176 --AVTNAEGILS-ATNPNPWPDTINQASQV-----NSAKSALNGDEKL 2217
QY 1843 AIGFO-AKADGEEAIVAIGRTOAGNOSIAIGDNAQATGQDSIAIGTG-----NVVAG 1893
Db 2218 AAKQTKASD-----IGRLTDLNNAQ-RTAANAEDVDQPNLAAVTAACKNATSLNTAMG 2270
QY 1894 KHSIGAIDPSTVKADNSYSGVNNQFTDATQTFVGVGNIIVTESNVALGSN---SAL 1950
Db 2271 NLKHALAEKDNTRKSVNY-----TDADQPKQQAQYDVTQAQEAITNANGSANETQV 2322
QY 1951 SAGTHAGTQAK-----KSDGTAGTPTTAGATGTVKGFA 1983
Db 2323 QRAALNQAKNDLNGDNKVAQAKETAKALASYNLNAQSTAASTQSDINATTVADVTA 2382
QY 1984 GQ-----TAVGAVSVGASGAEERRIONVAAGEVSATSTDAVNGSQLYKATQCIANATNE 2036
Db 2383 AQNTANELNTAMGOLONGIND-----QNTVKQOVNFTDADQ-GKKDAY--TNAVTAQGI 2434
QY 2037 LDHRIHONENKANAGISSAWAMASMPQAYIPGRSMYTG-----GIATH--NGQ---2082
Db 2435 LDRKANGQNTKAQ--VEAALNQVTTAKNALNGDANVROAKSDAKANLGTTLHLNNAQKOD 2492
QY 2083 -----GAVAV-GLSKLSDNGQ 2097
Db 2493 LTSQIEGATTVNGSVSKTKAQ 2514

RESULT 12
AB0480
probable invasin YPO3944 [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0480
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3013 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC93406.1; PID:gl15981852; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3944
```

```
Query Match 5.6%; Score 598; DB 2; Length 3013;
Best Local Similarity 21.2%; Pred. No. 1.8e-15;
Matches 547; Conservative 314; Mismatches 984; Indels 736; Gaps 118;

QY 13 TCTFWAVAYAKSHSTGGSCATGQVGSVRTLSFAR-----IAALAVLVIGATLNGSAYA 67
Db 519 THTISAVAYDAQGNASN-----RAVTSIEVTRPETWISHLATTIDNATANGIATN 569
QY 68 GIGISEADGGKGANARGDKSIAIGDIAQALGQSQSIAGDNKIVHNSNNNANIGAKASGN 127
Db 570 TVQATVTDGD-----GQPIIQGLINFVAVTQATLSTTEATG--ANGT 610
QY 128 ESIALGGDVLASGHASIAIGS-----DOLYLKKEVQOQISEL-LPIIRQKQKALNDIYOLA 181
Db 611 ASYTLTHTVSGVRSVVLGSSRSRVDVTFVADESTAEITAANLTVTTNDSVANG-----S 666
QY 182 DNLOKYRTHAQGHA-----STAVGAMSVAKGHFSNAFGTRATATAGTYSVLAVGLTA 233
Db 667 DTNVVRKVTDAYTNAVANSQVIFSASNGATVIDQTVITNAEG--IADST-----LTN 717
QY 234 TAKAASSIAGVSNQAIGAATAVGGSTQ-----VNL-NRGIALGFGS---OVLQKDNVDN 285
Db 718 TTAGSVWVTATLGGGSGQVDTTTPKPGSTAAISLVKLADRAVADGIDQNEIQVLRDGTGN 777
QY 286 AA-----NVRA-----YAPDDNQPIDNRYKATFXN--CATDVFSI-----318
Db 778 AVPNVPMISQADNGAIVASTENTGVDGTINATFTNLRAGESVSVSPALVGMTMTMTF 837
QY 319 -----GNSNGDS--IRRKIINVAGSADTDVAVNVAQLKEAVRLANR 358
Db 838 SADPRTAVVSTLAAIDNNAKADGDTNVVRVAVVDANGNSVPGSVTFPDAGNGAVLAQNP 897
QY 359 QITFKGDDSN-----RVEKGLGKLT---ITGGAQTSALTTHNIGVQVN 400
Db 898 VYTDNGRYAENTLTNLAIQTTTTCATTTVPQGVQVNTHFVAGAVDTITLTPVNGAVAN 957
QY 401 G---DGLKVLQAE-----LTSIKMV--TENTLTAKEKTVGKT-----RLTDDKIGFT 444
Db 958 GYNTSVQAVVSDSGNPGVTGATVVFSSNTAQVTTVIGTGVGDIATLINTVAGTS 1017
QY 445 NDMNGIDESKPYLDKDTGTHAGGQKITLTA---GVVDDAATYQOLKKVNOTAESALOT 501
Db 1018 NVVATIDTVNANI--DTAFVAGAVATITLTPAVNGAVADGADT---NQVDALVE-----1066
QY 502 FTVKKVDKNGN-----DANDSKLITVGKNNKPDG-----530
Db 1067 -----DANGNPITGAUVVVFSSANGATILSTMTNGVNGVASTLLTHTVAGTSNVVATVD 1120
QY 531 --TOVNTLKLKGNG-VDVTTETNGVTGFLNQNNGLTVGNSTLNN--DGLSVKNTNSN 584
Db 1121 TVNANIDTTFVAGAVATITLTPVNGAVADGANSVQAVVSDSDGNPVTGAUVVFSSAN 1180
QY 585 KOIQV-----GADGITFDISNKPAGIENNTTRITRDGIGFANNTGSLDANKPRLTP 637
Db 1181 ATAQITTVIGTTGADGIATATLTNTV--AGTSNVV-ATIDTVNANIDTAFVAGAVATITL 1237
QY 638 T-----GINAGGKELTN-----VQSALNPATNGQQLDFMNR-----LSTANTEKSSAAT 682
Db 1238 TAPVNGAVADGAD--TNQVDALVQDANGNAITGAUVVFSSANGADIATPMTNGVNGVAST 1296
QY 683 IKDLYNLSQVPLTFAGDGTGPNVTKKLGEILKVKGGKTATDADLTKNNGIVVADSDTNSLTV 742
Db 1297 L-----LHTVAG-----TSNVVATIDTISANIDTA 1322
QY 743 KLAKTLSLDLDAVNTKTLTASDKVTVDSGNNTAKLQ-----NGD-----LTFESKONTGA 790
Db 1323 FVA-----GAVATITLTPVNGAVADGADTNOVDALVEDANGNPITGAUVVFSSAN-CA 1375
QY 791 TPATNSKTIQVUGLKFET-----DNNGIALDGTYYITKDKVGFAGQDGLSKPYLD 842
Db 1376 TILSSTMTNGVNGVASTFLTHTVAGTSNVVATIGSVTENIDTAFVAGAVATITLTPV--1433
QY 843 KDKLVKVEVE--ITTINGINA-----CGKAITGLSNTLTDATNATTGHTVQLGIVDSTDXT 895
```


Db 1434 ----NGAVADVNTNSVQAVVSDSGNAVGTGATWFFSSA-NATAQITTVIG-TTGADGI 1486
QY 896 RAASIGDVLNAGELNKNKGDAKPVSTYDVTDFINGNATTAKVTYDGGKASKVADVNVVDG 955
Db 1487 ATATLTNTVAGTSNV-----VATIDV---NANIDTFFVA--GELENIV--VSIIN 1530
QY 956 TTIIHLTGADGNKNGIGVKTTLTKTDAGKDAKAINFVSNSGDDKA-LINAKDIADNLNTLA 1014
Db 1531 NNALANGADNIVEAFV-----TDRFGNGVANSGLIFGTNGASIVGSSSTVTNLD--- 1580
QY 1015 GEIR-----NTKGPADTAL-----QTFQVKVK-----ENGDDND 1045
Db 1581 GRVEASATHVAGSSNTVIAISGAHQGVYRTFVADVSTAGLKLTLFSLDNLQLANGKAGNI 1640
QY 1046 ADTITVCKDAKTQV--NTLKLKNGLDIQTNKGIVTFGI-----NTQSL----- 1091
Db 1641 AQALVT--DAHDNLLANQSVSFALDNGAVIESQGDASSAGIVLMRPNNTLAGMTTWTAT 1698
QY 1092 --KAGNNTLTNNGL-----SIKNTAGNEQIQVGADGVKFAKVNNGVVG- 1133
Db 1699 LDSTGQETLETHEFVAGKAASIEMTMTKDNVANNIDTNEVQLVLTVDVGNALINGAVNVL 1758
QY 1134 -----AGIDG--TTRITRDEIGFAGTNGSLDKSKPHLSKDGINA---GGKK 1174
Db 1759 TNSGMNITPNSVTTGSDGTATATLTHTLAGSLPINARIIDQ-----VSKT-INATFIADAS 1813
QY 1175 ITNIQSGEIAQNSNDVATGGKIYD-----LKTELENKISSTAKTAQNSLHEF 1221
Db 1814 TAILIAGDMFLVNDQVANGVAVNAQVAVTDSVGNFIKQTFVEFLSNNGTTIQYELDVT 1873
QY 1222 SVADEQGNFTVSNPYS-----SYDTSKTSDDVT-----FAGENG 1256
Db 1874 SV--EGGVMTFTTLAGITNTVATVSSGSSRNIDTTFIADVTTHAIAASDLMIIVDDA 1931
QY 1257 ITTKVNGVVRVIGIDQKLTTPKLTGVNNGKGIVIDSONGQNTITGLSNTLANVTNDK 1316
Db 1932 VADNLDKNEHARVTDKGNVLSQTVFTSGNGAALTIVNGISDGLTKATLTHTLAG 1991
QY 1317 GSVRTTEGKIIKDDEKTRAA--SIVDVLGAFNLQNGEADVPVSFYDT---VNFADGN 1371
Db 1992 TSVVTVARGNRVQSKDTTFIADRTTATIRASDLTITRNALADGVATNAARVITVDANG 2051
QY 1372 ATTAKVYDDTSKTSKVVDVNVDDTTFIEVKDKLGVKTTLTSTGTGANK-----FALS 1426
Db 2052 PVPSMEFVYSDNGALLTPTSGMTDSS-----GTFSTFTHTTAGISKVTAIIVTMG 2103
QY 1427 NQATGDALVKASDIVAHLNTLSGDIQTAKASQNSS-----AGYVDADGN-----KVI 1475
Db 2104 ISQTKDAVFIADRSTAHVS-----ELIVVKNDSLANNSDRNIVQAHIKDAHGNVVTGMNVN 2159
QY 1476 YDSTDNKKVYQAKNDGTVDKTKVAKDKLVAAQOTPDGTLQAMNVKSVINKEQVNDANKQ 1535
Db 2160 FSATENVTLTRANTVTT--NSQGYAENTLRHNAPVTSATV-----ATDLV 2204
QY 1536 GINEDNAFVKGL-----EKAASD-----NRTKNAAVTVGDLNVAQTPLTFFAGDTG 1581
Db 2205 GLTEDRVFVAGAGARIELFRINDGAVADGQTNRVEARVYDVSD-NLVPNSNVVFSADNG 2263
QY 1582 -----TTAKKLGLET-LTIKGGQDTNKLTDNNGIVVAGT-----DGTVKLAKD--- 1624
Db 2264 GQLVQNDVQTDALGSAVTVSNINTGVTKVTVTADGVSASTTTTFFIADRDRTATLTVDRFL 2323
QY 1625 LTN-----LNSVNAAGTKIDDKGVSFVDSGQAKANTPVLISANGLDLGGKV 1670
Db 2324 ITHDNVANGVVENRVLHLVDANDNSVGVNFSAITNGASINASAITINGFAIG--V 2381
QY 1671 ISNVKGTQKTDAAANVQQLNEVRNLGLGNAGNDN-----ADGNQVNIADIKKDPNS 1722
Db 2382 LTNTLSGFSDVTVTLVTP-----GGTESLTVTPQFIADINTANIA----- 2421
QY 1723 GSSSNRTVIAKTVILGGKGNNDTEKLATGVQGVGDKNANGDLSNVWK----- 1773

Db 2422 --TGDFVIIDDGAVANSVDANEVRVVT-----DNQGNAIAGYSVVFSSQNGATITTS 2472
QY 1774 --TQKQSGKALLATYNAAGTNYLTN-----NPAE 1802
Db 2473 GITVDGWSAKL-THIKAGESGILARLSRPMATVHTLMPYFIADVSTATLQLFNFNP-- 2529
QY 1803 AIDRINEQGIIRFF---HVNDSGNOEPVVGQNRNGIDSSAGKHSVAIGFQAKADGAAVAIG 1859
Db 2530 -IPIIADGVNQVFFVLGRVFDANQNFV--GGQQVAFSATNEVLTESNGSISTPESGVLLS 2586
QY 1860 -RQTAQGNQSIAGIDNAQATGDQSIAGTGNVAGKHSIGAIG-----DPSTVKADNYSV 1913
Db 2587 VTSTQAGVHPH-----TCTLSNNYTDTFGAFFIANKNTAQIUSTLMVV 2629
QY 1914 GNNQFTDATQTDVFGVGNNTVTTESNSVALGNSAISAGTHAGTOAKK---SD----- 1964
Db 2630 DNNALADGVTRNQVRA---HVDSTGNSVA---DMAVTFTANRGAQUSKVTVLTDNNGDA 2693
QY 1965 -----GTAGTTTTAGATGTVKGFAGQTAGAV--SVGASGAERRIQ- 2003
Db 2684 VNTLTSLVGVTVVTAKLGTAGTGPLTVDTVFTAGPLATLTLVTTVNNAFADNSATNTVQA 2743
QY 2004 -----NVAAGEVSATSDAVNGSOLYKATOGIANATNELDHRHONENKANAGISAM 2056
Db 2744 TLKDVSGNPTVGEVVAFA--ASNGATITATDGGVSNANG-----IVLATLNGTAGVSTVT 2797
QY 2057 A-----MASMPQAVIPGRSM-VTGGIATHNGQGAJA-----VGLSKLSDNGQWVFKINGS 2106
Db 2798 ATETLTETTTDTTFIAMKNLDVTVNGTTFNGDAGFTTGFVGAT-----FKVNSGG 2848
QY 2107 D 2107
Db 2849 D 2849

RESULT 13
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohca, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu-
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447

Query Match 5.5%; Score 590.5; DB 2; Length 2271;
Best Local Similarity 17.0%; Pred. No. 2.4e-15;
Matches 385; Conservative 435; Mismatches 973; Indels 469; Gaps 78;

QY 24 KSHSTGGSCATQGVSVRTLSPFARIAALAVLVIGATLN-----GSAYAGIG 70
Db 267 KQYMTTSGNATYQSTGIVTLTQDAYSQKGAITLGRIDSNKSFPHFSGKVNLGKRYEGHG 326
QY 71 ISEADGCGK-----CANARGDKSIAIGDIAQALGSQSIAGDNKIVHNSNNANI 120
Db 327 ----NGSDGIGFAPFPGVLGETGLNGAAVGIGLSNAFGFKL-----DYHNT-SKPS 375
QY 121 GAKAGNESIAIGDVLASGHASIAIGSDDLYLKKTVOQISELLPIRQKALNDIYQL 180
Db 376 AAKANADPSNVAGGAFAPVTTDSYGVATVTTSSSTADNAKL-----NVQPTNNTFQD 430
QY 181 ADTNLOKRYRTHAQGHASTAVGAMSVAKGHFSNAGFTRATAGTYSLAVGLTATAKAASS 240

Db 431 FDINY-----NGDTKMTVKYAGQWTRNISDWIAKSTTFNLSMTASTGCA-- 478
QY 241 IAVGSAQAALGFAA-----TAVCGSTQVNLNRGIAL-----GFGSOVLKQNDVNRAN 288
Db 479 ----TNLQOVQFOTFBYTESAVTQVRVYDVTGKDIIPPKTYSGNVQDVVITIDNQOASLT 534
QY 289 VRAYAPDDNQPIDNRYKATFKNGATDVFSIGNGNDSIRRR-KIINVAGSADTDVAVNA 347
Db 535 AKGY----NTSVDSSTASY-----NDTNKTVKMTNAG----- 564
QY 348 OLKEAVRLANRQITFKGDDSNRRVERKGLGKTLTITGGAQTSAL-----TDHNIG 396
Db 565 -----QSVTYVFTDV-----KAPTIVTGNQITIEVGKTMNPVWLTTDNGTG 605
QY 397 VYQNG--DGLKVLQALAEILTSKMTWTENLTANKEVTVGKTRLLTDDTKGFTNDMMGIDESKP 455
Db 606 TVTNTVTGLPSGLSYDSATNSIITGPTKIGQSTVTVVSTVQDANKKSTTTFTINVDVTTAP 665
QY 456 YL-----DKDTGIIHAGGQKITKLTAGVVDDDAATVYQLKKVNOTA 495
Db 666 TVTPIGQSQSEVYSPISPIKIATQDNSGNAVTVTVGLPSGLTFD-----STNNYI 716
QY 496 ES-----ALQFTVFKVKVKNQNDANDSKIIIVGKNNKPGDT-----QVNTLKILKG 540
Db 717 SGTPTNIGTSTISIVSTDASGNKTTTTPKYEVTRNGMSDSVSTSGTQQSQSVSTSKADS 776
QY 541 ENGVDVTTETGTV-----TFGLNQNLTVGNSTLANNGLSVKNTNSNKQIQVGADGIIF 596
Db 777 QS-----ASTSGSIVVSTASTSKTSVLSLSDSVASKSLSTSES-----VSSSTS 826
QY 597 TDISNKP-----GAGIENTRITRIGIGFANNTGSLDANKPRLTPTGINAGGKELTNVQ 651
Db 827 TSLVNSQSVSSMSGVSFKSTLS--DSISNSST-----EKSESLSSTSDSLRTSTSL 880
QY 652 SAINPATNGQLDFMRLTANTKESGAATIKDLYN---LSQVPITPAGDTGPNVTKKL 708
Db 881 DLSLMSST--GSLKSQSLSTISIGSSSTASLSDSSTNAISTSTLSSESASTSDSISIS- 938
QY 709 GEILKVGKGTTRADDITKINIGVVAUSTNSLTVKLAKLSDLDVAVN-----TKTITASD 763
Db 939 NSIANQSASTSKSDSQSTSL--STSDSKSMSTSESLSTSTSGSVSGSLIAASQ 995
QY 764 KTVTDSGN--NTAKLQGLDFTSKQNTGATPATNSKTIIGVDGLKFTDNNGIA-----LDGT 817
Db 996 SVSTSTSDSMSTSEIVSDSIS--GSLASDSKSMVSSMSMSTSGSSTSESLSDSQ 1051
QY 818 TYITKDKVGA---KQDGLDKSKPYLDKDKLVKVEVEITTINGINAGGKAITGLSNTLTD 874
Db 1052 STSDSDSKSLSLSTSQSGSTST--TGSVRTSESQSTSGMSASQSDSMGISTSFSD 1109
QY 875 ATNATTGHVTQLGIVDSTDKTRAASIGDVLNAGFNKNGDAKDFVSTVDTVDVDFINGNAT 934
Db 1110 ST-----SDSKASTASESISQASTSTSGS-----VSTSTSTSNSEBT 1151
QY 935 TAKVTYDGKASKVAYDVNVWDGTTIHLTGADGNKQIGVKTTLTKTDKADKKAINFV-- 992
Db 1152 STSVS-----DSTLSLSESDSISESTSTSDSISEALSASESTISISEN 1197
QY 993 NSGDDKALINAKDIADNLANTLAGEIRNKTGTADTALQTFQVKVKVKGDDNDADTTIVG 1052
Db 1198 STSDSESQASAFSLSESLSESTSESSESVSSTSESTSLSDSTSESGSTSTSLSNSTSG 1257
QY 1053 KDAKTQVNTLKLKGNGLDITQNKDGTVPFGINTOSGLKAGNNTILNNGLSIKRTAGN 1112
Db 1258 SASISTSTISE-----STSTFKSESVSTLSMSTSTSLSNSTSLSTSLSDSTSDSKS 1310
QY 1113 EQIQVG---ADGVKEKAVNGVVGAGIDGTTRITRDEIGFAGTNGSLDKSKPHLSKDGIN 1169
Db 1311 DLSLSTMSSTSDSISTSKSDSISTSTSLSGSTSESED-----STSSSESKSD----- 1357
QY 1170 AGGKKITNIQSGIAQNSDAVGGKIYDLKTELENKISSTAKTQNSLHEFSVADEQGN 1229

RESULT 14

AB3528

extracellular serine proteinase (EC 3.4.21.-) [imported] - Brucella melitensis (strain 1
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AB3528

Db 1358 -----STSMGISMOSQSTSGSTSTSTSLSDSTSTSLSLASMQSQGVDSNSASQASN 1411
QY 1230 NPTVS-----NPYSSVDTSKTSDDVITPAGENGITTKVNVGVVRVGDITQKGLTTPK 1280
Db 1412 STSTSTSESDSOSTSTYTTSQSTQSSESTSTSLSD--STSIKSTSQSGSTSTSLSGS 1470
QY 1281 LTVGNNGKGIVIDSQNGQNTITGLNLTANVTNDRKGSVRTTBOGKIIIKDEDKTRAASIV 1340
Db 1471 ESESDSQSISTASESTSESASTSLSDSTSTNSGSASTSTSLSNASASESDSSSTSL 1530
QY 1341 DVLASGNFNLQNGEAVDFVSTVDVNPADGNATKATVYDDTSKTSKVYVDVNVDDTTIE 1400
Db 1531 DSTSASMQSESSESQSTASLSLSTSTSNRMSTIASLSTSVSTSE--SGSTSESTSE 1587
QY 1401 VKDKKLGKVTTLTSTGTGANKFALSNQATGALVKASDIVAHLNLTSGDIQTAKGASQA 1460
Db 1588 SOSTSTSLSDSOSTSASTSASGASTSTSTSDSRSTASSTSTSMRTSTSDSQSMSTST 1647
QY 1461 NBSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTEKVAADKLVAQAQTPDGTLAQNVK 1520
Db 1648 STS-----MSDST--SLSDSVSDSTSDSTSTSGMSVSVSLSDSTSTSTAS 1694
QY 1521 SVIN-----KEOVNDANKKQGINEONAFVKGLEKAASDNKTNAAVTVGDLNAVAQ 1571
Db 1695 EVMSASISDSQSMSESVDNS--ESVSEGN-----SESDSKSMGSGSTSVSDSGSL- 1742
QY 1572 TPLTFAGDTGTTAKK---LGEITLIKGGQTDINKLTDNNIGVVAGTGDFTVKLAKDLTNL 1628
Db 1743 -----VSTSLRKSESVSSESLSGSQS---MSDS-----VSTSDSLSLSTSLRSS 1786
QY 1629 NSVNAAGTKIDDKGVSFVDSGQAKANTFVLSANGLDLGGKVISNVGKGTKDTDAANVQ 1688
Db 1787 ESVSSESDSLSDK-----STSGSTSTSTSGSLSTSTSLSGS--ESVSESTSLSDSLSMSD 1839
QY 1689 LNEVRNLGLGNAGNDNADGN-QVNIADIKQDONGSGSSNRVTVIKAGTVLGGKNNDTEK 1747
Db 1840 STSTSDSDSL--SGISLSGSLSTSLSDSLSDSKSLSSQS---MSGSESTSTSVSDSQS 1894
QY 1748 LATGGVQV-----GVDKQGNANGDLSNVVTKQDKGSKALLATYNAAGQNTYLNPA 1801
Db 1895 SETSNSQFDSMISASESDSMSTSDSNT---SGSNSTSTSTSLSTSDSMGSGSVSTST-- 1949
QY 1802 EADIRINEQIRFFHVNDGNQBPVQVRNGIDSSASGKHSAIGFOAKADGAAVAIGRQ 1861
Db 1950 -----SLSDS-----ISGSTSVSDSSSTSTSTSL----- 1973
QY 1862 TQAGNQSIAGDNAQATGQSIAGTGNVNVAGKSGAIGDPSTVKADNYSVGNNOQPTD 1921
Db 1974 --SDSMQSQSTSTASGSLSTSTSTSMGSASTS-----SSQSTSVSTSLSTSDSID 2025
QY 1922 AQTQDVFVGNNI-TVTESNVALGNSAISAGTHAGTQAKKSDGTAGTITTAGATGTVK 1980
Db 2026 STSISISGQSTVESESTSDSTSIDSESLSTSDSDSTSTSTSDSTSGSTSTSISSSLST 2085
QY 1981 GFAGQTAVG-----AVSVGASGAERRIONVAAGEVSATST-----DAVNGSQ-- 2022
Db 2086 SGSGSTSVSDSTSMSESDSTSVMSQDKSDSTSIDSESVSTSTSTSLSTSDSTSTSESL 2145
QY 2023 --LYKATQIANAT-----NEL-----DHRIHONENKANAGISS-AMAMASMP 2062
Db 2146 STSMGSGQISDSTSTSMGSGSTSTSESNMHPDSMSMHHTSTSTSLRSLSEATTSTSES 2205
QY 2063 QAYIPGRSMVTGGIATHNQGAVALGSLKSLSDNGQWVPKING 2104
Db 2206 QSTLSATSEVT---KHNGTTPAQSE--KRLPDTGDSI-KQNG 2240

R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, A.; Title, The genome sequence of the facultative intracellular pathogen *Brucella melitensis* A; Reference number: AD3252; PMID:11756688
A; Accession: AB3528
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-2554 <KUR>
A; Cross-references: GB:AB008918; PIDN:AAU53389.1; PID:gl7984282; GSPDB:GN00191
A; Experimental source: strain 16M
C; Genetics:
A; Gene: BMII0148
A; Map position: II
C; Keywords: hydrolase; serine proteinase

Query Match 53%; Score 568.5; DB 2; Length 2554;
Best Local Similarity 21.4%; Pred. No. 2.1e-14;
Matches 492; Conservative 271; Mismatches 925; Indels 607; Gaps 104;

202 GMSYAKGHFSAFGTRATAGTYSVLAVGLTATAKAASSIAVGSNAQAIGFAATA----- 256
20 GAIGFTK---AGAGSVLTGQNTYSGVTTLSGGILVTVMANAGVSGIG-OSTADPANL 75

257 -----VGGSTQVNLNRGIALFGSQVLQKDNVNAANVRAYAPDNPIDNRYK-- 305
76 MLESGTFRYTGGS--VITDRGFTLVNGGPVAVIEVTGSGSNL-AFSGIVTSPDDAGFEKK 132

306 ----ATFKNGATDVFSIGNSGNDSIRRKIIINVG-----AGSADTDVAVNAQLKBAVRU-- 355
133 GAGTLTFLNGSNHIGATTVTSSGGLTAVSLADGGQVSSGLKSGSDATNLILAGGALNYLG 192

356 ----ANRQITPKGDDSNRRKEGLKGLTITIGGQTSALTADHNGVTVQNGDGLKVLQALBT 411
193 STTSRDSFTLGGAGNSIGVANA-GTTLMSG--TAVGTD--GLTKLGGD----- 237

412 LTSKAMVTENLTANEKVTYVTKRLTTDKI----GFTNDMNGIDESKPYLKDGTGIIHAGG 467
238 --TLLSGTNYTGTAVNAGVLRAGSAQAFGPSGLMTVNG-----ASLELGG 284

468 QKIT--KLTAAGVDDDDAATYGQLKKVNAESALQPTPKVKKVQKNGNDANDSKIIIVGK 524
285 YDITVSGILGAGTVDLGNTL-----TSSGSAANSFTGKITGTGGFTRTGGSTQTLSG 337

525 NKKPDGQVNTLKLKGENVGVTTETNGTTFVFLGNQ--NGLTVGNSTLNDGLSVKNTN 582
338 CNS-DYTKTKTTIAGNGLSVCLNKGQAGSSNAPDLNVLNGLSTGTNTV-TTD 395

583 SNKIQVGADGITFTDISNKPAGAGIENTTTRITRDGIGFANNTGSL---DANKPRLTPTG 639
396 RGFTIQGGTGALSVTDAATLTFSG----QVVGFGALQKRDGTGLVLMNSNSYR-GGTS 449

640 INAGKELTNVQSAINPATNGQGLDFMRLSTANTEKSGSAATIKDL--YNLSQVPLTFA 697
450 VDAG-----TLRAGSGAFGGGMSLSN-----AAGAILDLGFDTSVTSLSGG 493

698 GDTGPNVTKKLEILKVGKGTATDLDL-----KNNIGV-----VADSTDNLSLT 742
494 GALGNVALG-GATLTISSGNSNGTSYTGALTGTGNFVKNNGTQRLTGCASSYSGSTTI 552

743 K-----LAKTSLDLDVNTKLTASDKVT-VDSGNNTAKLQNGDLTFKQNTGATPATNSK 797
553 NGGVLEVSLADGGSVSIGMSADADNLVINGGVLRVTGSGDSTDRQFTLGASGGNSIE 612

798 TIGVDGLKFTDNGI---ALDGTYYITKDKVPKQDGLSKPKPYLDKDKLKVEVEIT 854
613 SEGTAIILFTSNAAVTFAAANTAQTTLTLTLAGTNTDNERG-----AQLT 657

855 TNGINAGKALITGLSNT-----LTDATNATTHVTLQ-GLVDSTDK-----TRAASIGDV 903
658 NN-----GSGTSLTKTDTGTWFLNDSSTYTG-VTKINGGVLSVDKLANGVLASSIGAS 711

904 LNAGFNLNKNGDAKDFVSTYDVTDFINGNATTAKVTYDGGKASKVAYDVNVVDGTTIHLTGA 963

712 SSAASNL-----IIGNDSTLRYLGTGDTTDRILFTLASGLTYESSGS 753
964 DG-----NKNQIGVKTTLTKTDKAGDKAINFVSNGDDKALINAKDADNINLTLAGERN 1019
754 GAIVFTDTGOVALADNNQARTIALGGK-----NTGD-----NTLAGSIGD 793
1020 TKGTDATLQTFQVKVKENGDDDDADDTITVGDKAKTNQVNTLK--LKGKNGLDIQNK 1077
794 A-GTGKTTLA-----KNDDGTGVLGTGNTTGTGPNINKGLLKING-----GT 835
1078 DGTVTFGI-NTQSGLKAGNNNTLNNGLSLIKNTAGNEQIQVGADGVKFAKNNVNGVVGAG- 1135
836 TGSILTSDIVVTDGGLIFNRSDTLAVGGL-----ISGAGFVTQSGSGTTLTGCANSYTGATS 891
1136 -IDGTTTRITREIGFAG-----TNGSLDKSKPHLSKDGINAGGKKTINIQSEIAQNSDA 1190
892 VSAGTLLVNGDQSAATGQTSVANGSI-----LGGSGIIGNVVVTD---GALAPGSGA 942
1191 -----VTGGKIYDLKTELENKIS-----STAKTAQNSL-- 1218
943 GTLTINGSLALSAGSILSMQLGQAVAGGALNDL-IEVKGNLTLDGLTLDVAETAGSGYCP 1001
1219 -----HEFSVADEQGNFTVSNPYSSYDTSKTSDDVITFAG-----ENGITTKVNGK 1264
1002 GIYRLINVTGSLTDNGLDIGMLPNGAGAIQTAAGQVNLLAGGTNFNFWGDGVGPKFNSA 1061
1265 VVRVGIOTKGLTTPKLTGVNNV-----GKGIVIDSQNGQNT 1301
1062 V-----DGGNGTWMQSSGNMTDATGNINASYSDGAFATGTGTAGTVDIDSLGQVK 1114
1302 ITGLSNTLAN--VINDKGSVRTTEOGKIIKDEDKTRA-----ASIVDVLSAGFNLONGE 1354
1115 AEGQPAIDSAVAVTGDK--LELTGPQSTIRVGDGTAGAAVIATINSULTQNTOLE-KTD 1171
1355 AVDFVSTYDVTNFADG--NATTAKVTYDGT--SKTSKVYDVNVVDPTTIEVKDKKLGVK 1409
1172 AGTLVLT-GANSYTGTAINGGTIRISSDNLGVASSDISFDGALNTTANITADRAIIL 1230
1410 T-----TTLTSTGTGANKFALSNQATGDALVKASDIVAHLNLTLSGDIQTAKGASQ 1459
1231 TGAGTLTLDASTTLLSGPISGTCALTKSGGTLL--SGTAH-----TGTTTTAGTFLQ 1284
1460 ANSSAGYVDADGNKV-----IYDSTDNKKYQAKNDGTVDKTKB-----VAKDKLVAQA 1507
1285 IGRGTSSIDGNIVNNGALVDFDRAGTLAYTGSISGCTLTKNGSSLTITMTGTSTYTGCT 1344
1508 QTPDGTLAQM--NVKSVIN-----KEQVNDANKKQGINEDNAFVKGLEKAASDNKTKN 1558
1345 TVSAGTLALQAGGQIKGTASLTVDGGAEVLIDGSGSQFATGAGASVVG-----T 1393
1559 AAVTVGDLNVAQTPLTPLFAGDTGTAKKIGETLTIKGGQDTNKLTDNNIGV--VAGT-- 1614
1394 GTVTVRGGTASFDLSLTNSATGNS-----TITVAG--SGSQMTQTGTATFAGLAGTAT 1445
1615 ----DGFTV-----KLAKDLNLSNVNAGTK-----IDDKGVS 1644
1446 VDILDDGTMISSGASVFGGQLPMDATGQVTISAGSGQWTIANALYARRGSITVDDGGVV 1505
1645 FVDSGQAKANTPVLNANGLDLGGKVISNVGKGTQD--AANYQQLNEVRNLGLGNAG 1702
1506 TAGSAVIGYADTGINNPE-TDL---VVTGAGSRPETTGELAITNSAANAAGSITADGG 1561
1703 NDNADGNQVNIADIKDPNSGSSSNRTVIRKAGTVLGGKNNNDTEKLATGGVQGVGDKDN 1762
1562 VVKVGGGALAMPGNVNLINIGAAAGGSPAHAGTLDAG-----TVMVAGSQINFNHD-- 1614
1763 ANGDLSNVVKTKQDGGKCALLATYNAAGQTNLTNNPAEAIIDRINEGIRFFHVNDGNQ 1822
1615 ---DASTTFSATISGAS-----VSGSGATLLTGNNSYAGLTITVTAGSLYI---DGDQ 1663
1823 EFVVGNGRIDSSAGKHSV-----AIGFOAKADGEAAVAIGRQTAQNGQSTA-----IG 1872

Db 1664 -----SMATGLTIVNPGTGGTGTIGDVTVASGGAINPGSGMAPGTLNING 1712
QY 1873 DNAQATGD-QSIAIGTGNVAG-----KHSAGIDPSPVTKADNS 1910
Db 1713 DLTASGSGTQSFQGANIPGGPLNDLINVGBDLVLAGTLQVDTSAAGTMDPGIYRFPNY 1772
QY 1911 YSGNNNQFT-DATQTDVFE---GVGNNTITVTEGNSVAL-----GNSAISAG--- 1953
Db 1773 TGTLSQNAWTVNLPSPDFYQVTSVAQOINLVNTAGLALRFWDGADPNKNGKLEGNGI 1832
QY 1954 -----THAGT-QAKSDGTAGTTTATAGATGVKGFAGQATVAGVS----- 1992
Db 1833 WQAFSAPDNGNDWMTETGNTINAPFQDATEAVFTGEGTGTVDVDSKGAIVSGIQFVTDG 1892
QY 1993 -----VBSGAERIRQNVAG---EVSATSDAVNGSOLYKATQG----- 2029
Db 1893 YIVNGDAINLVGASGSTRVGGTGTGTVDVDSKGAIVSGIQFVTDG 1952
QY 2030 -----IANATNEL-----DRIHQENKANAG----- 2051
Db 1953 TGGTKITGGTLOVAKDSALGTRTCELLIDGGTLNTTADMTIDRSGTIDQAGTLDIDTGT 2012
QY 2052 --ISSAMAMSPQAYI---PGRSMVTGGIATHNGQGAVALSKLSD--NGQWVFKING 2104
Db 2013 LKIDGVLSCAG---AFVKTGAGRLAGDDHTYNGDASIASGTALTALGALGGMVNGIDG 2069
QY 2105 SADTQGHVGAACVAG 2119
Db 2070 RLEATGRVGAATTNSG 2084

RESULT 15
A97859
190-Kda cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: A97859
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 233, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; PMID:21442074; PMID:11557893
A:Accession: A97859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03811.1; PID:g15620410; GSPDB:GN00173
C:Genetics:
A:Gene: rompA

Query Match 5.1%; Score 550.5; DB 2; Length 2021;
Best Local Similarity 21.5%; Pred. No. 7.6e-14;
Matches 511; Conservative 243; Mismatches 811; Indels 811; Gaps 116;

QY 14 GTFMAVAYAKSHSTGGSCATQGVSRFLSPARIALAVLVIGA-----T 60
Db 88 GDYITADVADHIITAINVADTTPIG---LNTAQTVVGSIVTGNLLPVTITAGKSLT 143
QY 61 LNSAVAGTIGISEADGGKGGANARGDKSIAIGDIA-----QALGSQSIAGDNKIVHNSN 115
Db 144 LAG-----NNADAANHFQGPADPNYTGGLGNIALGGANAALIIOQAAPAKITLAGNIN 195
QY 116 NNANIGAKASGNESIAIGG-DVLASGHASIAIGSDLLYKKTETVQOISELLPIIRQKAL 174
Db 196 GGGIITVKDAAINGTIGNTNALATVNVGAGITLEGAIKATTKLTINAAVSL----- 249
QY 175 NDIYQLADTNLQYRRETHAQQGHASTAVGAMSAYAKHFSNAFGFR-----ATAEQTYSILAVG 230
Db 250 -----TLTNVNAVLTGALDNTTGVNDVNVGLNGLALGALSOVTG 285
QY 231 LTATAKAASIAVGSNAQAGFAATVAGSGTQVNLNRGIALFGSOVLQK---DNDVNAA 287
Db 286 NIGNTNALATISVAGKATILGGA--VIKATTKLTADNASAVTFTNPVWVGALDNTGNA- 342

QY 288 NVRAYAPDDNQPIDNRKYATFKNGATDVFSIGNSGNDSIRRKLIINVAGSAD----- 340
Db 343 -----NNGIVTFTGDSVTGTGNIGNTNA-----LATISVAGKATILGGAIIK 383
QY 341 -----TDAVNVAQLKEAVRLA-----NRQITFKGDSDNN-----RVEKG 374
Db 384 ATTKLTADNASAVTFTNPVVTGADTNGNANGIVTFTGDSVTGTGNIGNTNALATISVG 443
QY 375 LGKTLITIGG---AQTSALTDHNIQVQNGDGLKVQLAETLTSUKMT-----TENUTAN 426
Db 444 AGKA-TLGGAIKATTKLTADNASAV-----TFTNPVVTGADTNGNAN-N 488
QY 427 EKVTVGKTRLTDKIGFTNDMGIDSKPYLDKDTGHAG-----GQKITKLTAGVWDDA 482
Db 489 GI VFTFGDSVTGTGNIGNTNALAT-----SVGAKATILGGAIIKATTKLTADNA 537
QY 483 ATYGOLKKNQOT-----AESALQOTFT-VKKVDKNGNDANDSKIITVCKNKP-DGTQ 532
Db 538 SAVTFTNPVVTGADTNGNANGIVTFTGDSVTGTGNIGNTNALATISVGAKATILGGA 597
QY 533 VNTLKLKGENGVDTTETNGTVTFGL-----NONNGLT--VGNSTLNNDGLSVKNTNSNK 585
Db 598 IKATTKLTADNASAVTFTNPVVTGADTNGNANGIVTFTGDSVTGTGNIGNTNALATISVG 654
QY 586 QIOVGA-----DG-----ITFTDINSKPKGAGIENTTITRDTGIGFANN 624
Db 655 TVNVGAGIATLEGAVIKATTKLTADNASAVLTNVNAVLTGA-IDNTGTG--DNVGVNL 711
QY 625 TGSILD-----ANKPRLTPTGINAGGKELTNVQSAI-----NPAATNGGOL 663
Db 712 NGALSQVTGTGNIGNTNALATISVGAKATILGGAIVIKATTKLTADNASAVTFTNPVVTGAI 771
QY 664 DFMRLSTANTEKSGSAIYKDYLNLSQVPLTFAGDTGPNVTKKLGE-----I 711
Db 772 D-----NTGNA-NGGIA-----TFTGDS--TVTGNIGNTNALATISVGAGL 809
QY 712 LKVGKGTITADDLTKKNIGVADSTNSLTVKLAKTLSDLDVNTKLTASDKVTVDG- 770
Db 810 LRVOG-----GVKSNNTIN-----LTD-----NASAVTFTNPVVTGAI 843
QY 771 NNTAKLQNGDLTFKQKQ--TGATPATNS-KTIGVDGLKFTDNNNGIALDGTFTYTKDKVGF 827
Db 844 DNTGNANGIVTFTGDSVTGTGNIGNTNALATISVGAKATILGGAIIKATTKLTADNA-- 900
QY 828 AKQDGLSKSPYLDKDKLVGVEVITTINGINAGGKAITGLSNTLTDATNATTHVTLQ 887
Db 901 ----SAVFTNPV-----VVTGALDNTGNANGIV-----TFTGDSVTG 936
QY 888 IVDSTDKTRAASIGDVLNAGENLKNG--DAK--DFVSTYDVTDFINGNATTAKTAVTDGK 943
Db 937 NIGNTNALATVNVG-----AGVTLQAGGSLDANNIDF-GARSTLEF-NG-----PLDGG 983
QY 944 ASKVAYDVNVDDGTTIHLTG--ADGNKNQICVKTITLTDAKGDKAIFNSVNSGD----- 996
Db 984 GNAIPY-----YFKGAIANGNAILNVNTKLTAYHLTIGTVAEINIAGNLAID 1034
QY 997 ----DKALNADKIDANLNTLAGIERTKGTADTALQTFQVKVKENGDDDN---DADTI 1049
Db 1035 ASAGDVITILNAQDI--HPRAL-----DSALVLSNLTGVGVN-----NILLAALV 1077
QY 1050 TVGKDAKTQVNTLKLKGNGLDIOQNKDGTVTFGINTQSGLKAGN--NTTLNNGLSIK 1107
Db 1078 AFGVDEGTVVD-----GGVNGLINISNVAGAA-----RNIGDVGKFKFTLIYNAVIT 1128
QY 1108 NTAGNEIQVAGDGVKFAKVNNGVWVGAGIDGTTTITRDRIGFAGTNGSDKSKPHLSKDG 1167
Db 1129 D-----DYNLEGIQNVLIN--NADFTSSTAFNAGTIQINDATYTTIDANNGLN--- 1175
QY 1168 INAGGKKIINIQSGETAONS--ND-AVTGKGIYDVKTELENKISSAKTAQNSHEFSVA 1224
Db 1176 IPAGNIKFAHADAQLILQNSSGNDRITILGANIDPDNDDEGIVILNSVTA----- 1225
QY 1225 DEQGNFTVSNPYSSVDTSKTSDVITFFAGENGITTKVNGKVVGVGIDQTKLITPKLTVG 1284

Db 1226 ---GKLTIAAGKTFGGAKHLQDIV-FKGE-----GDFGTAGTTF----- 1261
Qy 1285 NNNGKGIVIDSQONITG---LSNTLANVTNDKGSVRITTEQKLIKDEKTRAASIVD 1341
Db 1262 --NTNIVLD-----ITGQLELGATTANVFLKDAVQLTQTG----- 1296
Qy 1342 VLSAGFNQNGEAVDFVSTYDVTNFDAGNATTAKVTYDDTSKTSKVYDVNVDDTTIEV 1401
Db 1297 -----NIGG-----FLDFNAKGTVTLLN-----WNVVAGT----- 1322
Qy 1402 KDKLGVKTTTLTSTGT-----GANKFALSNQATGDALVKASDIIVAHLNLTLSGDIQTAKGA 1457
Db 1323 -----VKNTGGTNGTLLVLGASNL---NRVNGIAMLKVG-----AGNVTIAKG- 1363
Qy 1458 SQANSSAGYVDADGNKYIYDSTDKYQAKNDGTVDKTKEVAKDKLVAQAQTPDGTLAQM 1517
Db 1364 --GNVKIGEIGTGTNTL---TLPAHF--KLTGSINKT-----GGQALKL 1401
Qy 1518 NVKSVINKEQVNDANKQGINEDNAFVKGLEKAAASDNKTNAAVTVGDLNAAVAQOTPLTFA 1577
Db 1402 N-----FMNG-----GSVSGVVGTAAANSV 1420
Qy 1578 GDTGTT-AKKLGEFLTITG-----GQTDNKLTDNNIGVVGAGTDGFTVKLAKDLT----- 1626
Db 1421 GDITTAGATSPASSVNAKGTATLGTTSTFAHTFTNTGAVTLAKGSITSFAKNVTATSEVA 1480
Qy 1627 NLNSVNAG-----GTKIDDKGV-----SFVDS-----SGQAKANTPVL 1659
Db 1481 NSATINFNSLAFNSITGSGTTLTGANQVYTYGTSGFTDTLNTTTFDGAAGSGGNIL 1540
Qy 1660 SANG--LDLGGKVTSNVKGKGTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIK 1717
Db 1541 IKSGSTLDSG--VSNLALVVTAT-----NFDMN----- 1567
Qy 1718 KDPNSGSSSNNRTVTKAGTVLGGK-----NND-----TEKLIATG 1751
Db 1568 ---NISPDTKVTISAEATAGGLKPTPKENVKITTINNDNRFPVDFPDASTLTLFAEDIAAG 1624
Qy 1752 GVQGVGDKDGNANGDLSNV-----WVKTKQDGSKK-----ALLATYNAAGOT 1793
Db 1625 -----VIDEDFAPGGPLANI PNAANIKKSLELMEDAPNGSDARQAFNFGMLTPIQEBADAT 1680
Qy 1794 NYLTNN---PAAIDIRINEQGI-----RPFHVNDGNOEPVVOGRNGIDSSAS 1837
Db 1681 THLMQDVVKPSDTIAAVNVQVVASNISNITALNARMKVKQAGNKGPVSSGDEDM----- 1736
Qy 1838 GKHSVAIGFQAKADGEAAVAIGRQTAQGNQSIATGDNAQATG-----DQSIATGTG 1888
Db 1737 -----AKFGAWISPFVGNATQKMCNSIS-GYKSDTTGGTIGDPGFVSDDIVLGLA 1785
Qy 1889 NVVJAGK---HSGAIGDPSTVKAD-----NSYSVGNNOFTDA----- 1922
Db 1786 YTRADTDIKLNKTKGDKRKNVESNIYSLYGLYVPYENLFVEAIIASYSDNKIRSKSRVI 1845
Qy 1923 ---TQTDVFGVGNNTIVTESNSVAL-----GSNSAISAGTHAGTQAKKSDGTAGT 1969
Db 1846 ATTLETGVGYQTANGVKSESYTGQLMAGVYWMSENINLTPLAGLRYSTIKDKSYKETGT 1905
Qy 1970 TTTAGATGTGKGFAGQTAGVAVSGVSGABERRIQNVAAGEVSAATSDAVNGSOLYKATQ- 2028
Db 1906 TY---QNLTVKGNKNTFDGLIGAKVS-----SNINYNEIVLT-----PELYAMVDY 1949
Qy 2029 GIANATNELDHRHQNENKANAGISSAMAMASMPQA 2064
Db 1950 AFKKVSAIDARLQ-----GMTAPLPTNSFKQS 1977

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:36:00 ; Search time 4.36825 Seconds
(without alignments)
3105.882 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGGKGANARGDKS.....GDIAQALGSQSIAGDNKIV 43

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	138	66.0	1964	2	Q8KQM9	Q8Kqm8 moraxella c
2	90	43.1	2314	2	Q8KQM8	Q8Kqm8 moraxella c
3	77.5	37.1	2190	16	Q8PGS0	Q8pgs0 xanthomonas
4	77	36.8	1299	16	Q9F3X6	Q9f3x6 pasteurella
5	77	36.8	2151	16	Q8PC05	Q8pc05 xanthomonas
6	72.5	34.7	1328	2	Q8LAX0	Q8lax0 xanthomonas
7	70.5	33.7	382	16	Q83J23	Q83j23 shigella fl
8	70.5	33.7	1588	16	Q8XD64	Q8xd64 escherichia
9	70	33.5	1953	16	Q8DHJ2	Q8dhj2 rhizobium l
10	69	33.0	762	16	Q8GIU6	Q8giu6 bradyrhizob
11	68	32.5	1461	16	Q8ZL64	Q8z164 salmonella
12	68	32.5	1778	16	Q8FCB2	Q8fcb2 escherichia
13	68	32.5	2712	16	Q9F3X5	Q9f3x5 pasteurella
14	67	32.1	1265	2	Q9FDA0	Q9fda0 xanthomonas
15	67	32.1	1291	16	Q92KQ7	Q92kq7 rhizobium m
16	66	31.6	1004	16	Q9PD63	Q9pd63 xylella fas

```

17 66 31.6 1309 16 Q8XPL8
18 65.5 31.3 641 16 Q8CKM1
19 65.5 31.3 658 16 Q8ZHU0
20 65.5 31.3 997 16 Q87DF4
21 65.5 31.3 1315 16 Q87D62
22 64.5 30.9 1114 16 Q8PGR8
23 64.5 30.9 1190 16 Q9PC04
24 64 30.6 372 16 Q89J37
25 64 30.6 1107 16 Q9F2D8
26 64 30.6 1116 16 Q83ST9
27 63.5 30.4 454 2 Q85267
28 63 30.1 261 11 Q9EQI3
29 63 30.1 487 2 Q9LA53
30 63 30.1 1264 2 Q8RQ61
31 63 30.1 2059 16 Q9PD50
32 62.5 29.9 436 16 Q9S2C3
33 61.5 29.4 148 10 Q8S608
34 61.5 29.4 190 16 Q8P769
35 61 29.2 369 16 Q89J38
36 61 29.2 539 2 Q8VSE9
37 61 29.2 550 2 Q9AJV8
38 61 29.2 567 2 Q9AFP8
39 61 29.2 615 16 Q05806
40 61 29.2 615 16 Q7TXQ0
41 61 29.2 663 16 Q8VJCO
42 61 29.2 694 16 Q53212
43 61 29.2 736 16 Q8CL86
44 60.5 28.9 251 16 Q829H6
45 60.5 28.9 544 5 Q9VMM5

```

ALIGNMENTS

```

RESULT 1
Q8KQM9 PRELIMINARY; PRT; 1964 AA.
ID Q8KQM9
AC Q8KQM9;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hemagglutinin.
OS HAG.
GN Moraxella catarrhalis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O35E;
RX MEDLINE=22112901; PubMed=12117964;
RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Geme J.W. III,
RA Hansen E.J.;
RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
RT Activities.";
RL Infect. Immun. 70:4523-4533(2002).
DR EMBL; AY077637; AAL78284.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 9.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 1964 AA; 201585 MW; 764A08F5F1F6854E CRC64;

```

Query Match 66.0%; Score 138; DB 2; Length 1964;

Best Local Similarity 69.8%; Pred. No. 2.4e-07;

Matches 30; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 IGISEADGGKGANARGDKSIAIGDIAQALGSQSIAGDNKIV 43

Db 69 IAIQSGGLNGARADGKESIAIGSIAQATGSGSIAGDNKAV 111

```

RESULT 2
Q8KQM8
ID Q8KQM8 PRELIMINARY; PRT; 2314 AA.
AC Q8KQM8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAG.
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1];
SEQUENCE FROM N.A.
RP STRAIN=O12E;
RX MEDLINE=22112901; PubMed=12117964;
RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Geme J.W. III,
RA Hansen E.J.;
RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
RT Activities."
RL Infect. Immun. 70:4523-4533(2002).
DR EMBL; AY077638; AAL78285.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 9.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 2314 AA; 238336 MW; 3PFD721F67D4E2FD CRC64;

Query Match 43.1%; Score 90; DB 2; Length 2314;
Best Local Similarity 52.3%; Pred. No. 0.13;
Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGGKGGANAR---GDKSIAIGDIAQALGQSIAIGDNKIV 43
DB 2040 AKADGEAAVAIGRQTQNGNSIAIGDIAQALGQSIAIGTGNV 2083

RESULT 3
Q8PGS0
ID Q8PGS0 PRELIMINARY; PRT; 2190 AA.
AC Q8PGS0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XAC3546
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1];
SEQUENCE FROM N.A.
RP STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo C.F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,

```

```

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AB012003; AAM38389.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 53.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2190 AA; 204715 MW; 1BEA011869A3C2CA CRC64;

Query Match 37.1%; Score 77.5; DB 16; Length 2190;
Best Local Similarity 45.2%; Pred. No. 3.6;
Matches 19; Conservative 7; Mismatches 11; Indels 5; Gaps 1;

QY 2 GISEADGGK-----GGANARGDKSIAIGDIAQALGQSIAIG 38
DB 1048 GESEAAQAQSTALGAAAGAYGDSGLAVGALSQAQSESTAMG 1089

RESULT 4
Q9F3X6
ID Q9F3X6 PRELIMINARY; PRT; 1299 AA.
AC Q9F3X6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mapa protein (Hsf).
GN MAPA OR HSF 2 OR PM1570.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1];
SEQUENCE FROM N.A.
RP STRAIN=Pm70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2];
SEQUENCE FROM N.A.
RP STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AJ277635; CAC14202.1; -.
DR EMBL; AB006194; AAK03654.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 10.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;

Query Match 36.8%; Score 77; DB 16; Length 1299;
Best Local Similarity 45.7%; Pred. No. 2.3;
Matches 21; Conservative 7; Mismatches 12; Indels 6; Gaps 2;

QY 1 IGISEADGGKG----GANARGDK--STAIGDIAQALGQSIAIGDN 40
DB 109 IGFGATNDGNTVAIGAKSKSAASTAIGDIAQALGQSIAIGDN 154

RESULT 5
Q8PCQ5
ID Q8PCQ5 PRELIMINARY; PRT; 2351 AA.

```


DR	EMBL; AF003000; BAB9874.1; -.
DR	GO; GO:0005622; C:intracellular; IEA.
DR	GO; GO:0005840; C:ribosome; IEA.
DR	GO; GO:0003735; P:structural constituent of ribosome; IEA.
DR	GO; GO:0006412; F:protein biosynthesis; IEA.
DR	InterPro; IPR008640; Hep Hag.
DR	InterPro; IPR008635; HIM.
DR	InterPro; IPR00911; Ribosomal_L11.
DR	InterPro; IPR005594; Yada.
DR	Pfam; PF05658; Hep Hag; 30.
DR	Pfam; PF05662; HIM; 8.
DR	Pfam; PF03895; Yada; 1.
DR	PROSITE; PS00359; RIBOSOMAL_L11; 1.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;
Query Match 33.5%; Score 70; DB 16; Length 1953;	
Best Local Similarity 35.4%; Pred. No. 24;	
Matches 17; Conservative 9; Mismatches 16; Indels 6; Gaps	
Qy	1 IGISADGGKG-----ANARGDKSTAIGDIAQLGSGSIAIGNKI 42 ::: : : :: : : : :
Dd	359 LGLGATAGGVDTGTAIGKOANASQDAIAMGTSAKASSAQIAVGINAV 406 ::: : : :: : : : :
RESULT 10	
Q89IU6	PRELIMINARY; PRT; 762 AA.
ID	Q89IU6
AC	Q89IU6
DT	01-JUN-2003 (TrEMBLrel. 24, Created)
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	BIR5538 protein.
GN	BIR5538.
OS	Bradyrhizobium japonicum.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC	Bradyrhizobiaceae; Bradyrhizobium.
OX	NCBI_TaxID=375;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=USDA 110;
RX	MEDLINE=22484998; PubMed=12597275;
RA	Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiomi T.,
RA	Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA	Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA	Tabata S.;
FT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT	Bradyrhizobium japonicum USDA110.";
RL	DNA Res. 9:189-197(2002).
DR	EMBL; AP005955; BAC50803.1; -.
DR	InterPro; IPR008640; Hep Hag.
DR	InterPro; IPR008635; HIM.
DR	Pfam; PF05658; Hep Hag; 12.
DR	Pfam; PF05662; HIM; 3.
KW	Complete proteome.
SQ	SEQUENCE 762 AA; 72776 MW; B2DFEE9A1CB241CE CRC64;
Query Match 33.0%; Score 69; DB 16; Length 762;	
Best Local Similarity 48.6%; Pred. No. 11;	
Matches 18; Conservative 4; Mismatches 15; Indels 0; Gaps	
Qy	4 SEADGGKANARGDKSTAIGDIAQLGSGSIAIGN 40 ::: : : :: : : : :
Dd	521 TKVNSAGAAANSGETDAIAGNGAQTQSIAIGLN 557 ::: : : :: : :~ : :
RESULT 11	
Q8ZL64	PRELIMINARY; PRT; 1461 AA.
ID	Q8ZL64
AC	Q8ZL64
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	BIR5538 protein.
GN	BIR5538.
OS	Bradyrhizobium japonicum.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC	Bradyrhizobiaceae; Bradyrhizobium.
OX	NCBI_TaxID=375;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=USDA 110;
RX	MEDLINE=22484998; PubMed=12597275;
RA	Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiomi T.,
RA	Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA	Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA	Tabata S.;
FT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT	Bradyrhizobium japonicum USDA110.";
RL	DNA Res. 9:189-197(2002).
DR	EMBL; AP005955; BAC50803.1; -.
DR	InterPro; IPR008640; Hep Hag.
DR	InterPro; IPR008635; HIM.
DR	Pfam; PF05658; Hep Hag; 12.
DR	Pfam; PF05662; HIM; 3.
KW	Complete proteome.
SQ	SEQUENCE 762 AA; 72776 MW; B2DFEE9A1CB241CE CRC64;


```

DE Putative inner membrane protein.
GN STM3691.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen K., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008871; AAL22550.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 10.
DR Pfam; PF05662; HIM; 13.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1461 AA; 147837 MW; 91C59A87E7282254 CRC64;

Query Match 32.5%; Score 68; DB 16; Length 1461;
Best Local Similarity 43.3%; Pred. No. 30;
Matches 13; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 13 ANARGDKSIAIGDIAQALGQSIAIGDNKI 42
DB 186 ATASGDDSAAFNGAKAIGTNSVALGSGV 215

RESULT 12
Q8FCB2 PRELIMINARY; PRT; 1778 AA.
AC Q8FCB2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin.
GN C4424.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016768; AAN82860.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 14.
DR Pfam; PF05662; HIM; 15.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1778 AA; 177678 MW; 98564AA3A797DA20 CRC64;

Query Match 32.5%; Score 68; DB 16; Length 1778;
Best Local Similarity 53.8%; Pred. No. 37;

DE Putative inner membrane protein.
GN STM3691.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen K., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008871; AAL22550.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 10.
DR Pfam; PF05662; HIM; 13.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1461 AA; 147837 MW; 91C59A87E7282254 CRC64;

Query Match 32.5%; Score 68; DB 16; Length 2712;
Best Local Similarity 57.7%; Pred. No. 59;
Matches 15; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 17 GSKSIAIGDIAQALGQSIAIGDNKI 42
DB 324 GESSVAIGDKAVSRGEASIAICKNAI 349

RESULT 14
Q9F3X5 PRELIMINARY; PRT; 1265 AA.
AC Q9F3X5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative outer membrane protein Xada.
GN XADA.
OS Xanthomonas oryzae (pv. oryzae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=64187;
RN [1] _
RP SEQUENCE FROM N.A.
RA Ray S.K., Rajeshwari R., Sonti R.V.;
RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
RT that is involved in virulence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288222; AAG01335.1; -.

```

```
DR InterPro: IPR008640; Hep_Hag.
DR InterPro: IPR008635; HIM_
DR InterPro: IPR005594; Yada.
DR Pfam: PF05658; Hep_Hag; 25.
DR Pfam: PF05662; HIM; 4.
DR DR Pfam: PF03895; Yada; 1.
SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;

Query Match 32.1%; Score 67; DB 2; Length 1265;
Best Local Similarity 46.2%; Pred. No. 34;
Matches 12; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 13 ANARGDKSIAIGDIAQALGSQSIAG 38
Db 694 AQTGVSVAIGELSKATGESVAVG 719

RESULT 15
Q92KQ7 PRELIMINARY; PRT; 1291 AA.
AC Q92KQ7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein R00458.
GN R00458 OR SMC01708.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rameperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591783; CAC41895.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR008640; Hep_Hag.
DR InterPro: IPR008635; HIM.
DR InterPro: IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 13.
DR Pfam; PF05662; HIM; 6.
DR Pfam; PF03895; Yada; 1.
DR PROSITE; PS00018; EF_HAND; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1291 AA; 127509 MW; 1B0F2A8CD1B1613C CRC64;

Query Match 32.1%; Score 67; DB 16; Length 1291;
Best Local Similarity 44.2%; Pred. No. 35;
Matches 19; Conservative 3; Mismatches 13; Indels 8; Gaps 1;

QY 4 SEADGGKGG-----ANARGDKSIAIGDIAQALGSQSIAG 38
Db 88 NEADPGGSDAIAIGTDAQNGDRSLAIGRQNGNEQSIGIG 130

Search completed: September 21, 2004, 23:48:29
Job time : 7.36825 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:32:39 ; Search time 0.799546 Seconds
(without alignments)
2800.358 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGKGANARGDKS.....GDIAQALGSQSTAIGDNKIV 43

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.5	31.3	434	1 YADA_YERPS	P10858 yersinia ps
2	63.5	30.4	455	1 YADA_YEREN	P31489 yersinia en
3	62.5	29.9	725	1 MASZ_SEPL	O05137 pseudomonas
4	61	29.2	377	1 RPOA_CHLMU	O93014 chlamydia m
5	60	28.7	278	1 T2D7_DROME	Q22272 drocephila
6	59.5	28.5	104	1 CH10_SCHPO	O59804 schizosacch
7	59	28.2	401	1 YK03_CAEEL	P34291 caenorhabdi
8	58	27.8	710	1 EFG_BUCBP	P59451 buchnera ap
9	57.5	27.5	403	1 CK05_HUMAN	Q9uhr6 homo sapien
10	57.5	27.5	676	1 ICP0_HSVBK	P29836 bovine herp
11	57	27.3	2946	1 NBEA_HUMAN	Q8afp9 homo sapien
12	56.5	27.0	245	1 YH77_ARCFU	Q28497 archaeglob
13	56	26.8	377	1 RPOA_CHLTR	P46449 chlamydia t
14	55.5	26.6	498	1 NDDO_ALCXX	P94212 alcaligenes
15	55.5	26.6	698	1 EFG_VTECH	Q9kuz7 vibrio chol
16	54.5	26.1	728	1 MASZ_BRUME	O8yir3 brucella me
17	54.5	26.1	728	1 MASZ_BRUSU	Q8fz50 brucella su
18	54	25.8	348	1 ID12_LACPL	Q89w06 lactobacill
19	54	25.8	1287	1 SK12_YEAST	P35207 saccharomyc
20	53.5	25.6	331	1 MACS_BOVIN	P12624 bos taurus
21	53.5	25.6	442	1 GLMD_RHIME	O87392 rhizobium m
22	53.5	25.6	699	1 EFG_HAEIN	P43925 haemophilus
23	53.5	25.6	699	1 EFG_VIBPA	Q87145 vibrio para
24	53.5	25.6	699	1 EFG_VIBVU	Q8dcq8 vibrio vuln
25	53.5	25.6	700	1 EFG_PASMU	P57938 pasteurella
26	53.5	25.6	703	1 EFG_ECOLI	P02996 escherichia
27	53.5	25.6	703	1 EFG_SALTY	P26229 salmonella
28	53.5	25.6	1043	1 DSG1_BOVIN	Q03763 bos taurus
29	53	25.4	219	1 PGNE_ECOLI	P77366 escherichia
30	53	25.4	507	1 FLIC_SALBE	Q06968 salmonella
31	53	25.4	507	1 FLIC_SALON	Q06974 salmonella
32	53	25.4	595	1 PRIM_CHLTR	O84799 chlamydia t
33	52.5	25.1	107	1 YBJQ_ECOLI	P75819 escherichia

RESULT 1

ID	YADA_YERPS	STANDARD;	PRT;	434 AA.
AC	P10858;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	01-JUL-1993 (Rel. 26, Last annotation update)			
DE	Invasin precursor (Outer membrane adhesin).			
GN	YADA OR YOPA OR INVA OR YOPI.			
OS	Yersinia pseudotuberculosis.			
OG	Plasmid pIB1.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Yersinia.			
OX	NCBI_TaxID=633;			
RN	[1]_TaxID=633;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=YPIII;			
RX	MEDLINE=88302441; PubMed=3043229;			
RA	Rosqvist R., Skurnik M., Wolf-Watz H.;			
RT	"Increased virulence of Yersinia pseudotuberculosis by two			
RT	independent mutations."			
RL	Nature 334:522-525 (1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=YPIII;			
RX	MEDLINE=89343638; PubMed=2761389;			
RA	Skurnik M., Wolf-Watz H.;			
RT	"Analysis of the yopA gene encoding the Yop1 virulence determinants			
RT	of Yersinia spp."			
RL	Mol. Microbiol. 3:517-529 (1989).			
CC	!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO			
CC	PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE			
CC	CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS			
CC	PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL			
CC	SURFACE.			
CC	!- SUBCELLULAR LOCATION: Outer membrane.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	X13883; CAA32088.1; -			
PIR	S04534; S04534.			
DR	InterPro; IPR008126; Adhesion.			
DR	InterPro; IPR008640; Hep Hag.			
DR	InterPro; IPR008635; HIM.			
DR	InterPro; IPR005594; Yada.			
DR	Pfam; PF05658; Hep Hag; 4.			
DR	Pfam; PF03895; Yada; 1.			
DR	PRINTS; PR01756; OMADHESIN.			
KW	Plasmid; Virulence; Signal; Outer membrane.			

P46723 mycobacteri
Q8ek71 shewanella
Q83887 coxiella bu
P57593 buchnera ap
Q8k948 buchnera ap
Q8zjb3 yersinia pe
Q88fi4 pseudomonas
Q8dh2 wiglesworth
Q8kam4 chlorobium
Q8fsv1 corynebacte
P24289 penicillium
P24504 penicillium

```

FT SIGNAL 1 25
PT CHAIN 26 434 INVASIN
SQ SEQUENCE 434 AA; 45054 MW; E2C55FB12B183D4 CRC64;

Query Match 31.3%; Score 65.5; DB 1; Length 434;
Best Local Similarity 46.5%; Pred. No. 2.4;
Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGGANARGDK--SIAIGDIAQ-----ALGSQSTAIIGNKI 42
| ||||| ||||| | : ||||| | :
| ||||| ||||| | : ||||| | :
Db 94 GAGGLNARAKDPYSIAIGATAEAAKPAAVAVGSGSIATGVNSV 136

RESULT 2
ID YADA_YEREN STANDARD; PRT; 455 AA.
AC P31489;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Invasin precursor (Outer membrane adhesin).
GN YADA OR YOPA OR INVA OR YOPL.
OS Versinia enterocolitica.
OG Plasmid pYV.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=630;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=6471/76 / Serotype O:3;
RA Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,
RA Skurnik M.;
RT "Hydrophobic domains affect the collagen-binding specificity and
RT surface polymerization as well as the virulence potential of the Yada
RT protein of Versinia enterocolitica.";
RL Mol. Microbiol. 10:995-1011(1993).
CC -|- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
CC SURFACE.
CC -|- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13882; CAA32086.1; -.
CC PIR; S04912; S04912.
CC InterPro; IPR008126; Adhesion.
CC InterPro; IPR008640; Hep Hag.
CC InterPro; IPR008635; HIM_
CC InterPro; IPR005594; Yada.
CC Pfam; PF05658; Hep Hag; 4.
CC Pfam; PF05662; HIM; 1.
CC Pfam; PF03895; Yada; 1.
CC PRINTS; PR01756; OMADHESIN.
CC PLASMID; Virulence; Signal; Outer membrane.
FT SIGNAL 1 25 INVASIN.
PT CHAIN 26 455
SQ SEQUENCE 455 AA; 47136 MW; AC12EF68C657DAC0 CRC64;

Query Match 30.4%; Score 63.5; DB 1; Length 455;
Best Local Similarity 44.2%; Pred. No. 4.3;
Matches 19; Conservative 6; Mismatches 9; Indels 9; Gaps 2;

QY 9 GKGG--ANARGDKSIATGDTAQ-----ALGSQSTAIIGNKI 42
| || ||| ||||| | : ||||| | :
| || ||| ||||| | : ||||| | :
Db 9 GKGG--ANARGDKSIATGDTAQ-----ALGSQSTAIIGNKI 42
| || ||| ||||| | : ||||| | :
| || ||| ||||| | : ||||| | :

```

```

Db 59 GAGGLNASAKGIHSIAIGATAEAAKGAAVAVGAGSIATGVNSV 101

RESULT 3
MASZ PSEFL
ID MASZ_PSEFL STANDARD; PRT; 725 AA.
AC O05137;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR FC2.4.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=L6.5;
RA Artgenave F.M., Delecu M., Vilagines R., Danglot C.;
RT "A functional glyoxylate bypass is mandatory for utilization of
RT alkanes by Pseudomonas fluorescens.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -|- PATHWAY: Glyoxylate bypass; second step.
CC -|- SUBUNIT: Monomer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y11998; CAA72726.1; -.
CC HSPSP; P37330; 1D8C.
CC HAWAP; MF_00641; -.
CC InterPro; IPR001465; Malate synthase.
CC InterPro; IPR006253; Malate synthG.
CC Pfam; PF01274; Malate synthase; 1.
CC TIGRFAMS; TIGR01345; malate syn G; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 725 AA; 78861 MW; 0C3326FB67C9B381 CRC64;

Query Match 29.9%; Score 62.5; DB 1; Length 725;
Best Local Similarity 40.7%; Pred. No. 8.8;
Matches 22; Conservative 2; Mismatches 17; Indels 13; Gaps 3;

QY 3 ISEADG---GKGGANARGDKSIAG-----DIAQALG-----SQSTAIIGNKIV 43
||||| ||| ||||| | : ||||| | :
| ||||| ||||| | : ||||| | :
Db 148 ISEADGAEGKGYKVRGDKVIAFAFAFLDEAAPLSAGSHVDSGTGYKIADGKLI 201

RESULT 4
RPOA CHLMU
ID RPOA_CHLMU STANDARD; PRT; 377 AA.
AC Q9PJN4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha
DE subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR TC0794.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiales; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]

```


FUNCTION: Eukaryotic CPN10 homolog which is essential for mitochondrial protein biogenesis, together with CPN60. Binds to

Best Local Similarity 42.4%; Pred. No. 12;
Matches 14; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 8 GKGGANARGDKSIAIGDIAALGQSQSIAGDN 40
DB 339 GGGSGGAPGSQVPMTEAQIIVQIAIRDN 371

RESULT 8

EFG_BUCBP
ID EFG_BUCBP STANDARD; PRT; 710 AA.
AC P59451;

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR BBP470.

OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;

RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola."
RL Proc. Natl Acad. Sci. U.S.A. 100:581-586(2003).

CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
the nascent protein chain from the A-site to the P-site of the
ribosome.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
EF-G/EF-2 subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AE014017; AAC27176.1; -;
DR HAMAP; MF 00054; -; 1.

DR InterPro; IPR0004540; EF-G.

DR InterPro; IPR000795; EF_GTPbind.

DR InterPro; IPR000840; EFG C.

DR InterPro; IPR009022; EFG III V.

DR InterPro; IPR005517; EFG IV.

DR InterPro; IPR004161; EFTU D2.

DR InterPro; IPR005225; Small GTP.

DR InterPro; IPR009000; Translat_factor.

DR Pfam; PF00679; EFG C; 1.

DR Pfam; PF03764; EFG IV; 1.

DR Pfam; PF00009; GTP_EFTU; 1.

DR Pfam; PF03144; GTP_EFTU D2; 1.

DR PRINTS; PR00315; ELONGATNFACT.

DR TIGRFAMs; TIGR00484; EF-G; 1.

DR TIGRFAMs; TIGR00231; small GTP; 1.

DR PROSITE; PS00301; EFATOR_GTP; 1.

KW Elongation factor; Protein biosynthesis; GTP-binding;
FT Complete proteome.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 710 AA; 79363 MW; 037BA37B190557FD CRC64;

Query Match 27.8%; Score 58; DB 1; Length 710;
Best Local Similarity 40.5%; Pred. No. 28;
Matches 15; Conservative 6; Mismatches 10; Indels 6; Gaps 2;

QY 13 ANARGD-KSIAIGDIAALGQSQSIAGD-----NKIV 43
DB 374 ANKREIKVRAGDIAAAIGLKSVTTGDTCLDPNNVV 410

RESULT 9

CK05_HUMAN
ID CK05_HUMAN STANDARD; PRT; 403 AA.
AC Q9UHR6;

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein C11orf5 (Protein FON).
GN C11ORF5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20069078; PubMed=10602999;

RA Lemmens I.H., Farnebo F., Piehl F., Merregaert J., Van de Ven W.J.M.,
RA Larsson C., Kas K.;
RT "Molecular characterization of human and murine c11orf5, a new member
of the FAUNA gene cluster."
RL Mamm. Genome 11:78-80(2000).

CC -!- TISSUE SPECIFICITY: Low expression in most tissues; highly
expressed in testis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF119497; AAF23591.1; -;
DR GenBank; HGNC:1177; C11orf5.

DR MIM; 604575; -;

DR InterPro; IPR007529; Znf HIT.

DR Pfam; PF04438; zf-HIT; 1

SQ SEQUENCE 403 AA; 42883 MW; 2384BF806CC40E71 CRC64;

Query Match 27.5%; Score 57.5; DB 1; Length 403;
Best Local Similarity 41.9%; Pred. No. 18;
Matches 13; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

QY 9 GKGGANARGDKSIAIGDIAALG---SQSIA 36
DB 296 GEGPTNQGYTLAALGDLAQTGLRARKQAVA 326

RESULT 10

ICP0_HSVBK
ID ICP0_HSVBK STANDARD; PRT; 676 AA.
AC P29836;

DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Trans-acting transcriptional protein ICP0 (P135 protein) (IER
2.9/ER2.6).

GN ICP0.

OS Bovine herpesvirus type 1 (strain K22).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.

OX NCBI_TaxID=31519;
RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92219360; PubMed=1313901;

RA With U.V., Fraefel C., Vogt B., Vlcek C., Paces V., Schwyzner M.;
RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
are 3' coterminal and encode a putative zinc finger transactivator


```

DR EMBL; AE004124; AAF93534.1; -.
DR PIR; C82332; C82332.
DR HSSP; P13551; 1ELO.
DR TIGR; VC0361; -.
DR HAMAP; MF_00054; -, 1.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR009022; EFG_III_V.
DR InterPro; IPR005517; EFG_IV.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF03764; EFG_IV; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP BIND 17 24 GTP (BY SIMILARITY).
FT NP BIND 88 92 GTP (BY SIMILARITY).
FT NP BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 698 AA; 76927 MW; F44FDA4DB8FC4ECC CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 698;
Best Local Similarity 46.4%; Pred. No. 53;
Matches 13; Conservative 4; Mismatches 10; Indels 1; Gaps 1;

QY 13 ANARGD-KSIAGDIAQAIGSQSIAIGD 39
Db ||| : ||| ||||| : |||
366 ANKRDEIKETIRAGDIAAAIGLKDVTGTD 393

Search completed: September 21, 2004, 23:43:19
Job time : 2.79955 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:24:53 ; Search time 4.42676 Seconds
(without alignments)
2744.569 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGGKGGANARGDKS.....GDIQAQIGSQSTAIGNKIV 43

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	209	100.0	43	2	Aaw32274 M. catarr
2	209	100.0	43	4	Aae00703 N-termina
3	209	100.0	2122	6	Abu08784 Moraxella
4	209	100.0	2123	4	Aae00701 Moraxella
5	190	90.9	40	4	Aae00705 N-termina
6	121	57.9	24	2	Aaw32276 M. catarr
7	121	57.9	24	2	Aae00702 Internal
8	90	43.1	1992	2	Aaw04505 Moraxella
9	90	43.1	1992	4	Aab69137 M. catarr
10	90	43.1	1992	4	Aab69133 M. catarr
11	90	43.1	2047	4	Aab69134 M. catarr
12	90	43.1	2053	4	Aab69135 M. catarr
13	90	43.1	2139	6	Abp71294 M. catarr
14	90	43.1	2314	4	Aab69136 M. catarr
15	89	42.6	1946	6	Abu35023 Protein e
16	81	38.8	3073	6	Abu21223 Protein e
17	76	36.4	2265	6	Abu17199 Protein e
18	76	36.4	2504	6	Ada34534 Acinetoba
19	70.5	33.7	1588	7	ADC01413 Enterohae
20	69.5	33.3	270	6	Abu19868 Protein e
21	69	33.0	1129	6	Abu20124 Protein e
22	68	32.5	1099	6	Abu20296 Protein e
23	68	32.5	1400	6	Abu45394 Protein e
24	68	32.5	1461	6	Abu47415 Protein e
25	68	32.5	1778	4	Abb52677 Escherich

ALIGNMENTS

RESULT 1
AAW32274
ID AAW32274 standard; peptide; 43 AA.

XX AAW32274;
AC AC
XX 08-MAY-1998 (first entry)
DT DT
XX DE
XX M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.
XX Outer membrane protein-106; OMP106; vaccine; immune response;
KW Cytotoxic antibody; Moraxella catarrhalis.
XX OS
XX Moraxella catarrhalis.

XX WO9741731-A1.
XX 13-NOV-1997.
XX 28-APR-1997; 97WO-US007679.
XX 03-MAY-1996; 96US-00642712.
XX (ANTE-) ANTEX BIOLOGICS INC.
XX Tucker K, Plosila L;
XX WPI; 1997-558601/51.

XX Outer membrane protein, OMP106, of Moraxella catarrhalis - used in vaccines for producing immune responses against M. catarrhalis.
XX Claim 9; Page 23; 78pp; English.
XX This is a peptide fragment of a novel outer membrane protein-106 (OMP106) of Moraxella catarrhalis. The OMP106 is an outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating cultivar. The peptide fragment can specifically bind to an antibody that binds the OMP106 polypeptide. The antibody is a cytotoxic antibody which mediates complement killing of M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be used in vaccines and antigenic compositions. They can also be used for producing an immune response in an animal against M. catarrhalis

SQ Sequence 43 AA;

Query Match 100.0%; Score 209; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.9e-19;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Abu39146 Protein e
Aaw56319 Haemophil
Abp76894 N. gonorr
Abu47474 Protein e
Aay13486 Mouse clo
Abu27363 Protein e
Abu34417 Protein e
Abu36862 Protein e
Abu36802 Protein e
Aaw56322 Haemophil
Abb64278 Drosophil
Aar56490 TATA-bind
Aaw06081 Drosophil
Aaw25025 TATA-bind
Abb62342 Drosophil
Aab27262 Arabidops
Aag40485 Arabidops
Aae25109 Arabidops
Aau72581 Arabidops
Aab27261 Arabidops

26 68 32.5 2712 6 ABU39146
27 67 32.1 2042 2 AAW56319
28 66 31.6 435 6 ABP76894
29 64 30.6 1107 6 ABU47474
30 63 30.1 261 2 AAY13486
31 61 29.2 394 6 ABU27363
32 61 29.2 505 6 ABU34417
33 61 29.2 615 6 ABU36862
34 61 29.2 594 6 ABU36802
35 61 29.2 2039 2 AAW56322
36 60.5 28.9 608 4 ABB64278
37 60 28.7 278 2 AAR56490
38 60 28.7 278 2 AAW06081
39 60 28.7 278 2 AAW25025
40 60 28.7 278 4 ABB62342
41 60 28.7 289 3 AAB27262
42 60 28.7 289 3 AAG40485
43 60 28.7 289 5 AAE25109
44 60 28.7 289 5 AAU72581
45 60 28.7 327 3 AAB27261

QY 1 IGISEADGGKGANARGDKSIAIGDIAQAALGSQSIATGDNKIV 43
 |||||
 Db 1 IGISEADGGKGANARGDKSIAIGDIAQAALGSQSIATGDNKIV 43

RESULT 2

AAE00703
 ID AAE00703 standard; peptide; 43 AA.

XX AAE00703;
 AC
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE N-terminal #1 of M. catarrhalis outer membrane protein-106 (OMP106).
 XX
 KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.

XX US6214981-B1.

XX 10-APR-2001.

XX 12-NOV-1997; 97US-00968685.

XX 03-MAY-1996; 96US-00642712.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L, Tillman UF;

XX WPI; 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.

XX Claim 2; Col 31; 49pp; English.

XX The present sequence is N-terminal of haemagglutinating Moraxella
 CC catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a
 CC therapeutic and prophylactic vaccine against M. catarrhalis infections of
 CC mammals. It is used for diagnosis of bacterial infections and as reagents
 CC for clinical or medical diagnosis of M. catarrhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarrhalis. It is also used as a probe to identify the
 CC presence of M. catarrhalis in biological specimens and to identify other
 CC bacteria that encode a polypeptide related to M. catarrhalis OMP106.
 CC OMP106-derived polypeptides are used as ligands to detect antibodies
 CC elicited in response to M. catarrhalis infections and also as immunogens
 CC for inducing M. catarrhalis-specific antibodies which are useful in
 CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis

XX Sequence 43 AA;

Query Match 100.0%; Score 209; DB 4; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1.9e-19;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISEADGGKGANARGDKSIAIGDIAQAALGSQSIATGDNKIV 43
 |||||

Db 1 IGISEADGGKGANARGDKSIAIGDIAQAALGSQSIATGDNKIV 43

RESULT 3

ABU08784
 ID ABU08784 standard; protein; 2122 AA.

XX
 AC ABU08784;

XX 28-MAY-2003 (first entry)
 XX Moraxella catarrhalis outer membrane protein, OMP106.
 XX
 KW Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;
 KW vaccine.
 XX
 OS Moraxella catarrhalis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..69 /label= Signal_sequence
 FT Protein 69..2122 /label= Mature_OMP106
 FT Region 69..111
 FT /label= N-terminal sequence of mature OMP106
 FT /note= "Specifically claimed in claim 9"
 FT Misc-difference 779..880
 FT /note= "Encoded by ATCTAAGGC"

XX US2002177200-A1.

XX 28-NOV-2002.

XX 20-MAR-2001; 2001US-00813214.

XX 03-MAY-1996; 96US-00642712.

XX 12-NOV-1997; 97US-00968685.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L;

XX WPI; 2003-328486/31.

XX N-PSDB; ABX93525.

XX Novel outer membrane polypeptide of Moraxella catarrhalis, useful for
 PT producing an immune response in an animal, and as ligands to detect
 PT antibodies elicited in response to Moraxella infections.

XX Claim 33; Page 28-34; 51pp; English.

XX The invention relates to an isolated or substantially pure outer membrane
 CC polypeptide of Moraxella catarrhalis (OMP106 polypeptide). The
 CC polypeptide and its peptide fragments are useful for producing an immune
 CC response in an animal and as ligands to detect antibodies elicited in
 CC response to M. catarrhalis infections. The polypeptide and its peptide
 CC fragment are further useful as active ingredients in vaccines against M.
 CC catarrhalis infections. The polypeptide is useful in immunoassays to detect
 CC an antibody binding the polypeptide is useful in passive immunisations
 CC M. catarrhalis in biological specimens, and also in passive immunisations
 CC against M. catarrhalis infections. An antibody binding the polypeptide is
 CC also useful to facilitate isolation and purification of the polypeptide
 CC and its peptide fragment, and as probes for identifying clones in
 CC expression libraries that have inserts encoding the polypeptide and
 CC peptide fragment. An antibody binding the polypeptide is also useful to
 CC diagnose M. catarrhalis infections. The polynucleotide encoding the
 CC polypeptide is useful as a probe to identify the presence of M.
 CC catarrhalis in biological specimens by hybridisation or PCR amplification
 CC and also to detect other bacteria that might encode a polypeptide related
 CC to M. catarrhalis OMP106. The polypeptide, an antibody binding the
 CC polypeptide, its peptide fragment and the polynucleotide encoding the
 CC polypeptide are useful as reagents for clinical medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis, as well as
 CC host defence mechanisms. The present sequence represents the amino acid
 CC sequence of M. catarrhalis outer membrane protein (OMP)-106

XX Sequence 2122 AA;

Query Match 100.0%; Score 209; DB 6; Length 2122;
 Best Local Similarity 100.0%; Pred. No. 1.6e-17;

XX WO9741731-A1.
 PN
 XX
 PD 13-NOV-1997.
 XX
 PF 28-APR-1997; 97WO-US007679.
 XX
 PR 03-MAY-1996; 96US-00642712.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L;
 XX
 XX WPI; 1997-558601/51.
 DR N-PSDB; AAT86522.
 XX
 XX Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 PT vaccines for producing immune responses against M. catarrhalis.
 XX
 PS Disclosure; Page 58; 78pp; English.
 XX
 CC This is a partial sequence of an amino terminal peptide fragment of a
 CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an
 CC outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating
 CC cultivar. The encoding DNA can be used as a 5' primer for PCR
 CC amplification of a full length OMP106 DNA. The DNA fragment can also be
 CC used as a probe for screening M. catarrhalis genomic libraries for OMP106
 CC polypeptide coding sequences. The encoded peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of M.
 CC catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis
 XX
 XX Sequence 24 AA;
 SQ

Query Match 57.9%; Score 121; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGGKGGANARGDKSIAIGDIAQ 28
 |||||
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24

RESULT 7
 AAE00702
 ID AAE00702 standard; peptide; 24 AA.
 XX
 AC AAE00702;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE Internal fragment of N-terminal outer membrane protein-106 (OMP106).
 XX
 KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.
 XX
 XX US6214981-B1.
 PN
 XX
 PD 10-APR-2001.
 XX
 PF 12-NOV-1997; 97US-00968685.
 XX
 XX 03-MAY-1996; 96US-00642712.
 PR
 XX (ANTE-) ANTEX BIOLOGICS INC.
 PA
 XX Tucker K, Plosila L, Tillman UF;
 PI
 XX

DR WPI; 2001-281002/29.
 DR N-PSDB; AAD04030.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 PS Example; Col 43-44; 49pp; English.
 XX
 CC The present sequence is the internal fragment of N-terminal
 CC haemagglutinating Moraxella catarrhalis outer membrane protein-106
 CC (OMP106). This sequence is used to design a probe and a 5' PCR primer.
 CC The OMP106 is used as a therapeutic and prophylactic vaccine against M.
 CC catarrhalis infections of mammals. It is used for diagnosis of bacterial
 CC infections and as reagents for clinical or medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also
 CC used as a probe to identify the presence of M. catarrhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used as
 CC ligands to detect antibodies elicited in response to M. catarrhalis
 CC infections and also as immunogens for inducing M. catarrhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarrhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarrhalis
 XX
 XX Sequence 24 AA;
 SQ

Query Match 57.9%; Score 121; DB 4; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGGKGGANARGDKSIAIGDIAQ 28
 |||||
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24

RESULT 8
 AAW04505
 ID AAW04505 standard; protein; 1992 AA.
 XX
 AC AAW04505;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-JAN-1997 (first entry)
 XX
 DE Moraxella 200 kDa outer membrane protein.
 XX
 KW Outer membrane protein; OMP; immunogen; vaccine; otitis media; diagnosis.
 XX
 OS Moraxella catarrhalis.
 XX
 XX WO9634960-A1.
 PN
 XX
 PD 07-NOV-1996.
 XX
 PF 29-APR-1996; 96WO-CA000264.
 XX
 PR 01-MAY-1995; 95US-00431718.
 PR 07-JUN-1995; 95US-00478370.
 PR 26-MAR-1996; 96US-00621944.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX
 XX Sasaki K, Harkness RE, Loosmore SM, Chong P, Klein MH;
 PI
 XX WPI; 1996-506162/50.
 DR N-PSDB; AAT38740.
 XX
 XX Moraxella outer membrane protein - useful as immunogen in protective
 PT vaccine and for diagnosis.
 XX
 XX Claim 14; Fig 6; 109pp; English.
 PS

XX An approx. 200 kDa outer membrane protein (AAW04505) can be isolated from
 CC Moraxella catarrhalis otitis media strain 4223 by electroelution, or
 CC expressed from a gene (see also AAT38740) obcd. from a strain 4223
 CC genomic library. Natural or recombinant outer membrane protein is useful
 CC as an immunogen to protect against infection by Moraxella, esp. M.
 CC catarrhalis. It can also be used to detect antibodies, esp. for
 CC differential diagnosis between bacteria that cause similar symptoms, and
 CC also useful as a carrier for other antigens and used to raise antitumour
 CC antibodies for conjugation to therapeutic agents. (Updated on 27-AUG-2003
 CC to correct OS field.)
 XX CC
 SQ Sequence 1992 AA;
 Query Match 43.1%; Score 90; DB 2; Length 1992;
 Best Local Similarity 52.3%; Pred. No. 0.024;
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
 QY 4 SEADGKGKANAR----GDKSTAIGDIAQALGQSQAIGDNKIV 43
 DB 1718 AKADGEAAVAIGRTQAGNQSTAIGDQAATGDSIAIGTGNV 1761
 RESULT 9
 AAB69137
 ID AAB69137 standard; protein; 1992 AA.
 XX
 AC AAB69137;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.
 XX
 KW Moraxella catarrhalis strain Q8; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; detection.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200107619-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-CA000870.
 XX
 PR 27-JUL-1999; 99US-00361619.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 XX
 DR WPI: 2001-159722/16.
 DR N-PSDB; AAF59106.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis.
 XX
 PS Claim 1; Fig 8A-V; 247pp; English.
 XX
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the M.
 CC catarrhalis M56 200kDa protein in pKS348, which is given in the

CC exemplification of the present invention
 XX
 SQ Sequence 1992 AA;
 Query Match 43.1%; Score 90; DB 4; Length 1992;
 Best Local Similarity 52.3%; Pred. No. 0.024;
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
 QY 4 SEADGKGKANAR----GDKSTAIGDIAQALGQSQAIGDNKIV 43
 DB 1718 AKADGEAAVAIGRTQAGNQSTAIGDQAATGDSIAIGTGNV 1761
 RESULT 10
 AAB69133
 ID AAB69133 standard; protein; 1992 AA.
 XX
 AC AAB69133;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.
 XX
 KW Moraxella catarrhalis strain 4223; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; detection.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200107619-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-CA000870.
 XX
 PR 27-JUL-1999; 99US-00361619.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 XX
 DR WPI: 2001-159722/16.
 DR N-PSDB; AAF59100, AAF59101.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis.
 XX
 PS Example 3; Fig 2A-W; 247pp; English.
 XX
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the M.
 CC catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is used
 CC in the exemplification of the present invention
 XX
 SQ Sequence 1992 AA;
 Query Match 43.1%; Score 90; DB 4; Length 1992;
 Best Local Similarity 52.3%; Pred. No. 0.024;
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
 QY 4 SEADGKGKANAR----GDKSTAIGDIAQALGQSQAIGDNKIV 43
 DB 1718 AKADGEAAVAIGRTQAGNQSTAIGDQAATGDSIAIGTGNV 1761


```

FT      /note= "signal peptide"
FT      67..2139
FT      /note= "mature protein"
FT      764..913
FT      /note= "specifically claimed immunogenic fragment"
FT      962..1200
FT      /note= "specifically claimed immunogenic fragment"
XX      WO2003004651-A1.
XX      16-JAN-2003.
XX
XX      01-JUL-2002; 2002WO-SR001299.
XX
XX      04-JUL-2001; 2001SE-00002410.
XX
XX      (FORS/) FORSGREN A.
XX
XX      Forsgren A, Riesbeck K, Janson H;
XX      WPI; 2003-221598/21.
XX      N-PSDB; AB258968.
XX
XX      New surface exposed immunoglobulin D-binding protein from Moraxella
XX      catarrhalis, useful for treating an autoimmune disease or as vaccine,
XX      comprises a molecular weight of 200 kDa.
XX
XX      Claim 1; Fig 6; 98pp; English.
XX
XX      The invention relates to a surface exposed immunoglobulin D-binding
XX      protein detected in Moraxella catarrhalis. The protein has an apparent
XX      molecular weight of 200 kDa, and can selectively bind membrane bound or
XX      soluble IgD. The protein or its variant or fragment, is useful in
XX      treating an autoimmune disease or as vaccine. The present sequence
XX      represents a M. catarrhalis the surface exposed protein (mid) of the
XX      invention
XX
XX      SQ      Sequence 2139 AA;
XX
XX      Query Match      43.1%; Score 90; DB 6; Length 2139;
XX      Best Local Similarity 52.3%; Pred. No. 0.026;
XX      Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
XX
XX      QY      4 SEADGGKGGANAR----GDKSIAIGDIAQAALGSQSIAGNDKIV 43
XX      :::||| | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      1865 AKADGEAAVAIGRQTQAGNQSIAGDQAATGDSIAIGTGNVV 1908
XX
XX      RESULT 14
XX      AAB69136
XX      ID      AAB69136 standard; protein; 2314 AA.
XX
XX      AC      AAB69136;
XX
XX      DT      24-APR-2001 (first entry)
XX
XX      DE      M. catarrhalis lesl 200kDa protein SEQ ID NO:11.
XX
XX      KW      Moraxella catarrhalis strain Q8; major outer membrane protein;
XX      KW      200kDa outer membrane protein; antibacterial; immunogenic; infection;
XX      KW      otitis media; detection.
XX
XX      OS      Moraxella catarrhalis.
XX
XX      PN      WO200107619-A1.
XX
XX      PD      01-FEB-2001.
XX
XX      PF      26-JUL-2000; 2000WO-CA000870.
XX
XX      PR      27-JUL-1999; 99US-00361619.
XX
XX      PA      (CONN-) CONNAUGHT LAB LTD.
XX
XX      Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX      WPI; 2001-159722/16.
XX      N-PSDB; AAF59105.
XX
XX      New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX      useful in protective vaccines and for diagnosis.
XX
XX      Claim 1; Fig 5A-Y; 247pp; English.
XX
XX      The present invention describes an isolated and purified nucleic acid (I)
XX      that encodes a 200 kDa outer membrane protein (II) has antibacterial activity and
XX      The 200 kDa outer membrane protein (II) has antibacterial activity and
XX      can be used in vaccines. (II), and its truncated versions, are used as
XX      immunogenic compositions and vaccines to protect against M. catarrhalis
XX      infections, particularly otitis media in humans. (II) is also used as
XX      antigen in immunoassays for detecting specific antibodies (Ab), and to
XX      generate Ab. (I) are used for recombinant production of (II) and its
XX      fragments are used as probes for identifying/cloning 200 kDa protein
XX      genes from other strains, and for diagnostic detection of M. catarrhalis.
XX      (I) makes possible production of large amount of recombinant immunogens.
XX      Expression of truncated versions of (II) reduces toxicity of the protein
XX      towards the Escherichia coli host. The present sequence represents the M.
XX      catarrhalis lesl 200kDa protein, which is given in the exemplification of
XX      the present invention
XX
XX      SQ      Sequence 2314 AA;
XX
XX      Query Match      43.1%; Score 90; DB 4; Length 2314;
XX      Best Local Similarity 52.3%; Pred. No. 0.029;
XX      Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
XX
XX      QY      4 SEADGGKGGANAR----GDKSIAIGDIAQAALGSQSIAGNDKIV 43
XX      :::||| | | | | | | | | | | | | | | | | | | | | | |
XX      Db      2040 AKADGEAAVAIGRQTQAGNQSIAGDQAATGDSIAIGTGNVV 2083
XX
XX      RESULT 15
XX      ABU35023
XX      ID      ABU35023 standard; protein; 1946 AA.
XX
XX      AC      ABU35023;
XX
XX      DT      19-JUN-2003 (first entry)
XX
XX      DE      Protein encoded by Prokaryotic essential gene #20550.
XX
XX      KW      Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX      OS      Moraxella catarrhalis.
XX
XX      PN      WO200277183-A2.
XX
XX      PD      03-OCT-2002.
XX
XX      PF      21-MAR-2002; 2002WO-US009107.
XX
XX      PR      21-MAR-2001; 2001US-00815242.
XX      PR      06-SEP-2001; 2001US-00948993.
XX      PR      25-OCT-2001; 2001US-0342923P.
XX      PR      08-FEB-2002; 2002US-00072851.
XX      PR      06-MAR-2002; 2002US-0362699P.
XX
XX      PA      (ELIT-) ELITRA PHARM INC.
XX
XX      PI      Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX      PI      Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX      DR      WPI; 2003-029926/02.
XX      DR      N-PSDB; ACA38893.
XX
XX      PT      New antisense nucleic acids, useful for identifying proteins or screening

```

PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX
XX Claim 25; SEQ ID NO 62947; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1946 AA;

Query Match 42.6%; Score 89; DB 6; Length 1946;
Best Local Similarity 52.3%; Pred. No. 0.032;
Matches 23; Conservative 4; Mismatches 13; Indels 4; Gaps 1;

Qy 4 SEADGGKGGANAR----GDKSIAIGDIAQAALGQSQSIAGGNKIV 43
Db 1672 AKADGEAAVAIGRTQAGSQSIAGDKAQAATGDSIAIGTGNVV 1715

Search completed: September 21, 2004, 23:39:48
Job time : 7.42676 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:33:49 ; Search time 1.42358 Seconds
(without alignments)
2905.511 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGGKGANARGDKS.....GDIQAALGSQSTAIGDNKIV 43

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	70.5	33.7	1588	2	A86036	probable adhesin Z
2	70.5	33.7	1588	2	H91188	probable adhesin E
3	66	31.6	1004	2	C82672	surface-exposed ou
4	65.5	31.3	434	2	S04534	invasin precursor
5	65.5	31.3	658	2	AH0110	probable surface p
6	64.5	30.9	1190	2	A82615	surface protein XF
7	64	30.6	1107	2	AC0976	probable autotrans
8	63.5	30.4	455	2	S04912	yopa protein - yer
9	63	30.1	2059	2	D82671	surface protein XF
10	62.5	29.9	436	2	T36478	probable solute-bi
11	61	29.2	394	2	B81663	DNA-directed RNA p
12	61	29.2	615	2	H70589	hypothetical glyci
13	61	29.2	694	2	F70868	hypothetical glyci
14	60	28.7	278	2	A49067	transcription init
15	60	28.7	327	2	T00797	hypothetical prote
16	59.5	28.5	104	2	T41381	Chaperonins 10 Kd
17	59.5	28.5	365	2	AB3486	cell surface prote
18	59	28.2	401	2	C88571	protein C05B5.3 (i
19	58.5	28.0	562	2	T49788	related to merozoi
20	58	27.8	293	2	AB3639	serine proteinase
21	58	27.8	1176	2	T18042	ice nucleation pro
22	57.5	27.5	676	1	EDBE22	immediate-early pr
23	57.5	27.5	1660	2	A70869	hypothetical glyci
24	57	27.3	364	2	AE0169	probable exported
25	57	27.3	422	2	S04911	yopa protein - yer
26	57	27.3	505	2	AC3486	cell surface prote
27	57	27.3	1737	2	AS9235	unconventional myo
28	56.5	27.0	245	2	H69471	conserved hypothet
29	56.5	27.0	2342	2	T18200	fatty-acid synthas

30	56	26.8	377	2	A71505	DNA-directed RNA p
31	56	26.8	552	2	T23755	hypothetical prote
32	56	26.8	584	2	G70804	hypothetical glyci
33	55.5	26.6	698	2	C82332	translation elonga
34	55	26.3	229	2	AB0220	flagellar L-ring p
35	55	26.3	447	2	E83465	conserved hypothet
36	54.5	26.1	297	2	AF2605	oxidoreductase Atu
37	54.5	26.1	297	2	E97387	hypothetical prote
38	54.5	26.1	340	2	T20807	hypothetical prote
39	54.5	26.1	467	2	F96773	geranylgeranyl red
40	54.5	26.1	562	2	B70953	hypothetical glyci
41	54.5	26.1	644	2	T15652	hypothetical prote
42	54.5	26.1	728	2	AF3299	malate synthase (E
43	54.5	26.1	741	2	G70317	hypothetical glyci
44	54.5	26.1	1538	2	H70846	hypothetical glyci
45	54	25.8	196	2	G85435	TINY-like protein

ALIGNMENTS

RESULT 1

A86036

Probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A86036

R:Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A86036

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1588 <STO>

A:Cross-references: GB:AE005174; NID:gi2518349; PIDN:AAG58749.1; GSPDB:GN00145; UNGP:Z50

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z5029

Query Match 33.7%; Score 70.5; DB 2; Length 1588;

Best Local Similarity 32.8%; Pred. No. 3.4;

Matches 19; Conservative 9; Mismatches 11; Indels 19; Gaps 2;

QY 4 SEADG-----GKG-----GANARGDKSIAIGDIAQALGSQSIAGDNKI 42

Db 281 NKADGVDAIALNGSQSRGLNTIALGTASNATGDKSLAUGSSNANGINSVALGDSI 338

RESULT 2

H91188

Probable adhesin ECs4480 [similarity] - Escherichia coli (strain O157:H7, substrain RIMD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
C:Accession: H91188

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H91188

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1588 <HAY>

A:Cross-references: PIDN:BA837903.1; PID:gi3363955; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs4480

Query Match 33.7%; Score 70.5; DB 2; Length 1588;

Best Local Similarity 32.8%; Pred. No. 3.4;

Matches 19; Conservative 9; Mismatches 11; Indels 19; Gaps 2;

```
C;Genetics:
A;Gene: yopa
A;Genome: plasmid

Query Match      31.3%; Score 65.5; DB 2; Length 434;
Best Local Similarity 46.5%; Pred. No. 3.6;
Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGGANARGDK--SIAIGDIAQ-----ALGQSQSIAGDNKI 42
Db 94 GAGGLNARAKDPYSIALGATREAKPAVAVGSGSIATGNSV 136

RESULT 3
C82672
surface-exposed outer membrane protein XF1516 [imported] - Xylella fastidiosa (strain 9a)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: C82672
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: C82672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1004 <SIM>
A;Cross-references: GB:AE003981; GB:AE003849; NID:g9106543; PIDN:AAF84325.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.F.G.; Nunes, L.R.; Oliveira, M.A.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1516

Query Match      31.6%; Score 66; DB 2; Length 1004;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 15; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 9 GKGGANARGDKSIAIGDIAQALGQSQSIAG 38
Db 576 GSGVASAIGKDPSTATGASQAQVGDSSVALG 605

RESULT 4
S04534
invasin precursor - Yersinia pseudotuberculosis plasmid pTBI
C;Species: Yersinia pseudotuberculosis
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S04534; S04910
R;Roqvist, R.; Skurnik, M.; Wolf-Watz, H.
Nature 334, 522-525, 1988
A;Title: Increased virulence of Yersinia pseudotuberculosis by two independent mutations
A;Reference number: S04534; MUID:88302441; PMID:3043229
A;Accession: S04534
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <ROS>
A;Cross-references: EMBL:X12758; EMBL:X13883; NID:g48639; PIDN:CAA32088.1; PID:g48640
R;Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A;Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
A;Reference number: S04910; MUID:89343638; PMID:2761389
A;Accession: S04910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <SKU>
A;Cross-references: EMBL:X12758; EMBL:X13883; NID:g48639; PIDN:CAA32088.1; PID:g48640
```

```
C;Genetics:
A;Gene: yopa
A;Genome: plasmid

Query Match      31.3%; Score 65.5; DB 2; Length 434;
Best Local Similarity 46.5%; Pred. No. 3.6;
Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGGANARGDK--SIAIGDIAQ-----ALGQSQSIAGDNKI 42
Db 94 GAGGLNARAKDPYSIALGATREAKPAVAVGSGSIATGNSV 136

RESULT 5
AH0110
probable surface protein (partial) YPO0902 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AH0110
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0110
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-658 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:g15978974; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0902

Query Match      31.3%; Score 65.5; DB 2; Length 658;
Best Local Similarity 39.0%; Pred. No. 5.5;
Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 2 GISEADGGKGGANARGDKSIAIGDIAQALGQSQSIAGDNKI 42
Db 480 GANSATCGAGSV-ASGNNSTAFSGKAKAANAALGANSV 519

RESULT 6
A82615
surface protein XF1981 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82615
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82615
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1190 <SIM>
A;Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN00
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
```

RESULT 9
D82671
surface protein xFls39 [imported] - *Xylella fastidiosa* (strain 9a5c)

```

RESULT 11
B81663      DNA-directed RNA polymerase, alpha chain TC0794 [imported] - Chlamydia muridarum (strain
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
C:Accession: B81663

```


M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: E84736
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-327 <STO>
 A:Cross-references: GB:AE002093; NID:G2914702; PIDN:AAC04492.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g32710; F24L7.15
 A:Map position: 2
 A:Introns: 193/2
 C:Superfamily: Arabidopsis thaliana hypothetical protein F24L7.15

Query Match 28.7%; Score 60; DB 2; Length 327;
 Best Local Similarity 33.3%; Pred. No. 12;
 Matches 12; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY	2	GISEADGGKGGANARGDKSIAGDIAQALGQSQIAI	37
		: :	
Db	12	GAGAGAGGGGGGGGGESSIALMDVVSPSSSLGV	47

Search completed: September 21, 2004, 23:44:38
 Job time : 3.42358 secs

